

*ARMY RESEARCH LABORATORY*



## **Quantal Response: Estimation and Inference**

**by Joseph C Collins**

**ARL-TR-7088**

**September 2014**

## **NOTICES**

### **Disclaimers**

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

# **Army Research Laboratory**

Aberdeen Proving Ground, MD 21005-5068

---

**ARL-TR-7088**

**September 2014**

---

## **Quantal Response: Estimation and Inference**

**Joseph C Collins**

**Survivability/Lethality Analysis Directorate, ARL**

<b>REPORT DOCUMENTATION PAGE</b>				<i>Form Approved OMB No. 0704-0188</i>
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. <b>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</b>				
<b>1. REPORT DATE (DD-MM-YYYY)</b> September 2014	<b>2. REPORT TYPE</b> Final	<b>3. DATES COVERED (From - To)</b> October 2013 – July 2014		
<b>4. TITLE AND SUBTITLE</b> Quantal Response: Estimation and Inference			<b>5a. CONTRACT NUMBER</b>	<b>5b. GRANT NUMBER</b>
			<b>5c. PROGRAM ELEMENT NUMBER</b>	
<b>6. AUTHOR(S)</b> Joseph C Collins			<b>5d. PROJECT NUMBER</b>	<b>5e. TASK NUMBER</b>
			<b>5f. WORK UNIT NUMBER</b>	
<b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b> U.S. Army Research Laboratory ATTN: RDRL-SLB-D Aberdeen Proving Ground, MD 21005-5068			<b>8. PERFORMING ORGANIZATION REPORT NUMBER</b> ARL-TR-7088	
<b>9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b>			<b>10. SPONSOR/MONITOR'S ACRONYM(S)</b>	
			<b>11. SPONSOR/MONITOR'S REPORT NUMBER(S)</b>	
<b>12. DISTRIBUTION/AVAILABILITY STATEMENT</b> Approved for public release; distribution is unlimited.				
<b>13. SUPPLEMENTARY NOTES</b> Author's Email: <joseph.c.collins38.civ@mail.mil>				
<b>14. ABSTRACT</b> Quantal response modeling denotes sensitivity testing or analysis of a binary response to a continuous stimulus. Historically, this has been computed from first principles by assuming a single simple parameterization and imposing the form of a specific cumulative distribution function on the response. Application of the Generalized Linear Model approach admits arbitrary response functions and model complexity in a unified framework and subsumes the historical approach as a specific case. The Likelihood Ratio approach to quantal response modeling allows statistical analysis of data with or without a zone of mixed results on equal ground. Confidence interval (estimation) and hypothesis test (inference) computations are documented with particular attention to the 2-parameter model most commonly used for V <sub>50</sub> ballistic limit armor acceptance testing. It is the aim of this paper to present the relevant theory and methodology in an open and complete manner and thus encourage discussion, criticism, and implementation by the larger community.				
<b>15. SUBJECT TERMS</b> ballistics, V <sub>50</sub> test, quantal response, confidence intervals, hypothesis tests				
<b>16. SECURITY CLASSIFICATION OF:</b> Unclassified		<b>17. LIMITATION OF ABSTRACT</b> UU	<b>18. NUMBER OF PAGES</b> 78	<b>19a. NAME OF RESPONSIBLE PERSON</b> Joseph C Collins
<b>a. REPORT</b> Unclassified	<b>b. ABSTRACT</b> Unclassified	<b>c. THIS PAGE</b> Unclassified		<b>19b. TELEPHONE NUMBER (Include area code)</b> 410-278-6832

---

## Contents

---

<b>List of Figures</b>	<b>v</b>
<b>1. Introduction</b>	<b>1</b>
1.1 Background .....	1
1.2 The Model.....	5
1.3 Properties of Maximum Likelihood Estimators .....	7
1.3.1 Confidence Interval .....	9
1.3.2 CI-based Hypothesis Test.....	10
1.3.3 Wald Test .....	10
1.3.4 Wald (CR) CIs on Mean Response .....	14
1.4 Suggested Approach .....	18
<b>2. Likelihood Ratio Estimation and Inference</b>	<b>19</b>
2.1 Likelihood.....	19
2.2 Likelihood Ratio Tests .....	19
2.3 Statistic Distribution .....	21
2.3.1 Asymptotics .....	21
2.4 Hypothesis Test Construction .....	21
2.5 Single Point .....	22
2.5.1 Confidence Region .....	22
2.5.2 Confidence Interval .....	22
<b>3. Application to Quantal Response</b>	<b>23</b>

3.1	The Quantal Response Model .....	23
3.2	Estimation .....	24
3.2.1	With a Zone of Mixed Results.....	24
3.2.2	With No Zone of Mixed Results.....	26
3.3	One-Sample Inference .....	27
3.3.1	Continuity Correction .....	29
3.3.2	Implementation.....	29
3.3.3	Likelihood Ratio Confidence Intervals.....	32
3.3.4	Confidence Interval Examples.....	34
3.4	Two-Sample Inference .....	54
3.4.1	G Test: the Basic Test .....	54
3.4.2	M Test: the Test of $X(0.5)$ .....	54
<b>4.</b>	<b>One-Sided Tests</b>	<b>63</b>
<b>5.</b>	<b>Computation</b>	<b>63</b>
5.1	Full Model.....	64
5.2	Null Model: $b = 0$ .....	64
5.3	Restricted Model: $G(x_o) = p_o$ .....	64
<b>6.</b>	<b>Conclusions</b>	<b>65</b>
<b>7.</b>	<b>References</b>	<b>66</b>
<b>List of Symbols, Abbreviations, and Acronyms</b>		<b>69</b>
<b>Distribution List</b>		<b>70</b>

---

## List of Figures

---

Fig. 1	Small-sample and asymptotic test power .....	14
Fig. 2	CR and LR CIs, various $\gamma$ .....	16
Fig. 3	Linearized CR and LR CIs, various $\gamma$ .....	17
Fig. 4	Zmr data .....	25
Fig. 5	Gap data .....	27
Fig. 6	Example of $P(x_o) = p_o$ with zmr data .....	31
Fig. 7	Example of $P(x_o) = p_o$ with gap data.....	32
Fig. 8	Multiple $P(x_o) = p_o$ with zmr .....	35
Fig. 9	LR CIs on $P(x)$ for zmr data.....	36
Fig. 10	LR CIs on $P(x)$ for gap data .....	37
Fig. 11	LR CIs on $X(p)$ for zmr data .....	38
Fig. 12	LR CIs on $X(p)$ for gap data .....	39
Fig. 13	LR CIs on $P(x)$ and $X(p)$ for zmr data.....	40
Fig. 14	LR CIs on $P(x)$ and $X(p)$ for gap data .....	41
Fig. 15	LR CIs bands for zmr data.....	42
Fig. 16	LR CI bands for gap data .....	43
Fig. 17	LR CI bands for gap data, continuity correction.....	44
Fig. 18	LR CI bands for gap data, continuity correction (CC) effect .....	45
Fig. 19	CR CIs on $P(x)$ for zmr data .....	46
Fig. 20	CR CIs on $X(p)$ for zmr data .....	47

Fig. 21 CR CIs on $P(x)$ and $X(p)$ for zmr data .....	48
Fig. 22 LR and CR CIs for zmr data .....	49
Fig. 23 LR CIs for Webb data .....	50
Fig. 24 CR CIs for Webb data.....	51
Fig. 25 LR and CR CIs for Webb data.....	52
Fig. 26 CR and LR CIs, $\gamma = 0.99, n = 16$ .....	53
Fig. 27 Two-sample tests, 2 zmr data sets, example 1 .....	57
Fig. 28 Two-sample tests, 2 zmr data sets, example 2 .....	58
Fig. 29 Two-sample tests, 1 zmr and 1 gap data set, example 1 .....	59
Fig. 30 Two-sample tests, 1 zmr and 1 gap data set, example 2 .....	60
Fig. 31 Two-sample tests, 1 zmr and 1 gap data set, example 3 .....	61
Fig. 32 Two-sample tests, 2 gap data sets .....	62

---

## **1. Introduction**

---

The ballistic limit test is an important part of characterizing armor performance or establishing the lethality of projectiles. At a minimum, the analysis of such a test provides a single number which is an estimate of the ballistic limit itself. By 1950, however, it was already established that statistically meaningful decision support requires additional methodology that manifests as confidence intervals and hypothesis tests on the ballistic limit.

Unfortunately, as time passed, this knowledge was ignored. In many important applications, the methodology reverted to obtaining only that single number as the product of a ballistic limit analysis. The purpose of this report is to resurrect the original ideas, generalized and cast into a modern framework, and thus to provide today's analyst with the tools required for proper decision support.

It is illuminating to begin by tracing the evolution (or devolution) of ballistic limit analysis through the literature.

### **1.1 Background**

Golub and Grubbs<sup>1</sup> wrote in 1950:

In current tests of armor plate, AP projectiles are fired at various velocities against a given plate until some mixture of complete penetrations and partial penetrations is obtained. The “Ballistic Limit”\* of the plate is then estimated somewhat arbitrarily by using the average velocity of two projectiles – one resulting in a complete penetration and the other a partial penetration. It is seen, however, that such a “rule of thumb” does not lead to a standard error which may be attached to the estimate of ballistic limit or the standard deviation of the distribution of velocities for the “zone of mixed results”. The “zone of mixed results” is an interval extending from the velocity for which the probability of penetration would be zero to a velocity at which the proportion of penetrations would be unity. The purpose of the present note is therefore to investigate a method for estimating the mean or median (50%) velocity, the standard deviation of the distribution of velocities for the zone of mixed results and to approximate standard errors or estimates of precision which one can be attached to the above two figures. The general problem here involves the analysis of so-called “sensitivity data”.

... In the execution of a test for determining the Ballistic Limit of a given armor plate, therefore, the results of such a test usually yield a set of distinct projectile velocities or testing levels. Assuming that each projectile has a “critical velocity” at or above which the projectile will penetrate the plate and below which it will fail to penetrate, and also that these critical velocities follow the Normal Distribution Law, then the general problem may be to determine for any given velocity the probability of a penetration. However, it is usually sufficient to determine that velocity for which the probability of

penetration is one-half (defined hereafter as the Ballistic Limit velocity). The remainder of this note is concerned with how the mean or Ballistic Limit velocity and standard deviation of the normally distributed critical velocities may be estimated from a unique set of velocities, and how the precision of our estimates may be approximated.

---

\*Ballistic Limit has, in many instances, been defined somewhat loosely as “that velocity as which a given type of projectile will penetrate a given (thickness and type of) armor plate”. However, it turns out that for a given series of AP projectiles fired at an appropriate constant velocity part of the projectiles will completely penetrate the plate and the remainder will not, the proportion penetrating depending on the level of velocity. Thus, for problems of this type it becomes necessary to regard the “Ballistic Limit” as a parameter of the probability distribution involving the proportion of successes for various levels of velocity.

In the next generation, the 1983 Engineering Design Handbook<sup>2</sup> and McKaig and Thomas<sup>3</sup> present references and FORTRAN implementations of the 1972 DiDonato and Jarnagin<sup>4</sup> method of estimation for the normal response model. The later 1984 US Army Developmental Test Command (USADTC) Test Operations Procedure (TOP) 2-2-710, Ballistic Tests of Armor Materials<sup>5</sup> cites an even earlier 1970 work with code by Hagan.<sup>6</sup>

The 1984 TOP states:

5.1.1  $V_{50}$  Ballistic Limit. . . If enough rounds are fired, two parameters, the mean and standard deviation, can be determined for each ballistic test; they are referred to as the  $V_{50}$  ballistic limit and the standard deviation, both expressed in meters per second. The standard deviation is a measure of the data spread or the steepness of the curve.

The 1984 TOP details 4 methods for ballistic limit analysis. The TOP presents the Up-and-Down method for normal distributions in which ballistic limit estimates “commonly” computed as averages of firing velocities.

5.1.1.1 Up-and-Down Method (for Normal Distributions) . . . The following varieties of the up-and-down method are commonly used in determining the  $V_{50}$  ballistic limit of armor:

- a. One complete penetration and one partial penetration within a velocity spread of 15 m/s . . . These two striking velocities are then averaged to obtain the ballistic limit.
- b. Two complete penetrations and two partial penetrations . . .
- c. Three complete penetrations and three partial penetrations . . . These six striking velocities are then averaged to estimate the ballistic limit.
- d. Five complete penetrations and five partial penetrations . . . These 10 striking velocities are then averaged to estimate the  $V_{50}$  ballistic limit.

The second TOP recommendation is the Langlie method, along with a 1970 computer program for calculation of parameter estimates when the data has a zone of mixed results (zmr). Otherwise,  $V_{50}$  is estimated as the average of 2 velocities.

#### 5.1.1.2 Langlie Method (for Normal Distributions) . . .

- f. If the firing does not produce a zone of mixed results, compute  $V_{50}$  by averaging the lowest complete and highest partial.
- g. If the firing produces a zone of mixed results, compute  $V_{50}$  and standard deviation by using the cumulative normal and the principle of maximum likelihood. A computer program is available at Aberdeen Proving Ground (APG) for this purpose (ref. 10e, Appendix J).

The third method estimates penetration probability, not ballistic limit.

#### 5.1.1.3 Sampling-of-Levels Method (Distribution Not Normal) . . . A point estimate of the probability of penetration is computed at each velocity level by determining the ratio of complete penetrations to the number of rounds fired.

The TOP explanation of the probit design, the final method, may be interpreted as advocating the appropriate computation, but this is not explicit.

#### 5.1.1.4 Probit Design (for Normal Distributions)

The probit design of test involves a number of trials at each of several preset levels of severity, and as such is similar to the sampling-of-levels method. The difference is that the term "probit design" is referred to in the literature as applying only to normal distributions; the sampling-of-levels method (a test devised at APG) is used for distributions that are not normal.

Occurrences of the terms "normal model" and "standard deviation" in the TOP indicate awareness that some technology other than arithmetic average exists, but no effort is made to explain its necessity or utility.

More recent guidance is in the 1997 MIL-STD-662, Department of Defense Test Method Standard,  $V_{50}$  Ballistic Test for Armor<sup>7</sup>:

3.7 Ballistic limit . . . Certain approaches lead to approximation of the  $V_{50}$  Point, that is, the velocity at which complete penetration and incomplete penetration are equally likely to occur.

3.8 Ballistic limit, protection criteria ( $V_{50}$  BL(P)). The  $V_{50}$  BL(P) may be defined as the average of an equal number of highest partial penetration velocities and the lowest complete penetration velocities which occur within a specified velocity spread. The normal up-and-down firing procedure is used.

5.4 Calculation of the  $V_{50}$  BL(P) ballistic limit. The  $V_{50}$  BL(P) shall be calculated by taking the arithmetic mean of an equal number of the highest partial and the lowest complete penetration impact velocities within the allowable velocity span as defined by the contracting officer (see USATECOM TOP 2-2-710).

5.7 Acceptance and rejection. The selected armor samples shall meet the minimum  $V_{50}$  BL(P) ballistic requirements specified in the order for the represented lot to be acceptable.

MIL-STD-662 optionally defines ballistic limit as an average of projectile velocities and unconditionally mandates calculation of the ballistic limit as such.

Various current detail specifications for lightweight high hard<sup>8</sup> and ultra high hard<sup>9</sup> steel armor applications specify  $V_{50}$  protection ballistic limit [BL(P)] testing. Both Detail Specifications (DTLs) ignore the correct computations and state:

A.2.6  $V_{50}$  protection ballistic limit, BL(P). The protection  $V_{50}$  ballistic limit is defined as the average of 6 fair impact velocities comprising the three lowest velocities resulting in complete penetration and the three highest velocities resulting in partial penetration.

Furthermore, both DTLs state:

A.4.1 Ballistic tests.  $V_{50}$  ballistic tests are to be conducted according to USADTC TOP 2-2-710, Ballistic Tests of Armor Materials or ITOP 2-2-713, Ballistic Testing of Armor to determine compliance with the requirements of Tables A-I through A-V.

The DTLs provide ballistic limit specifications for various threats, obliquities, armor thicknesses, and armor classes. The TOP provides some elements of test protocol and data collection such as sample sizes, required numbers of complete and partial penetrations, admissible velocity ranges, and determination of fair hits. The TOP references data reduction through computation of empirical ballistic limits as averages of firing velocities or, optionally, via probit regression parameters. But the TOP does not document how the tests are to be conducted. Neither do the DTLs. The actual conduct of the test is implemented as a decision rule for comparing the data to the specification. In the absence of such guidance, the inexperienced analyst will apply the naive rule:

accept the sample if the empirical  $V_{50}$  exceeds the specification value. (1)

It will be seen that this incurs a Type I error of  $\alpha = 0.5$ , which is equivalent to discarding the data and flipping a coin to determine the outcome of the test.

In 1950, Golub and Grubbs<sup>1</sup> criticized the “rule of thumb” calculation of  $V_{50}$  as an average of firing velocities and pointed out that parameter estimates and parameter variance estimates are required for error bounds (equivalently, for hypothesis testing). They present maximum likelihood estimation (MLE) computations and solution via paper and pencil, thus providing the asymptotic technology needed to calculate error bounds and conduct the required statistical hypothesis tests. The TOP (almost 35 years later) mentions such computations but does not illustrate or mandate their use. In spite of the 1950 admonition, MIL-STD-662 (almost 50 years later) and the referenced DTLs (almost 60 years later) all implement the naïve procedure and go so far as to actually define  $V_{50}$  as an arithmetic average of velocities.

There is a backwards progression in methodology from Golub and Grubbs in 1950 through the TOP in 1984, the MIL-STD in 1997, and the DTLs in 2008 and 2009 to the pre-1950 situation. Statistical decision support is not required in the MIL-STD, TOP, and DTLs.

The standards go through great detail describing various experimental designs, which are nothing more than procedures for selecting firing velocities and sample size. Statistical decision support (inference) is another matter entirely.

The various 2&2, 3&3, 4&4, etc.,  $V_{50}$  estimates obtained without a zmr are not useful for inference in the asymptotic MLE framework. Any decision based on comparing such a point estimate to a specification value is devoid of statistical properties and cannot be justified.

Even in the presence of a zmr, the literature is ambiguous (i.e., confused) about the meaning of “standard deviation”, whether it refers to the variance of the critical velocity distribution or the variance of the ballistic limit estimate, and completely silent about its utility.

## 1.2 The Model

One approach to the development of the sensitivity model uses a distribution of tolerances (critical velocities in the ballistic setting). Assume that the random unknown and unobservable critical velocity  $C$  is distributed with probability law (cumulative distribution function, cdf)

$$\Pr[C \leq v] = F_C(v). \quad (2)$$

Upon firing a projectile, one observes its velocity  $v$  and penetration  $y \in \{0, 1\}$ . If  $C \leq v$ , then velocity exceeds critical velocity, penetration occurs, and  $y = 1$ . Otherwise  $y = 0$ . Thus,

penetration is random with probability  $p$  where

$$p = \Pr[y = 1 \mid v] = F_C(v). \quad (3)$$

The usual parameterization of  $F_C$  is location-scale, with parameter vector  $\theta = (m, s)$ ,

$$F_C(v) = G\left(\frac{v - m}{s}\right). \quad (4)$$

Common choices for  $G$  include the normal law

$$G(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-u^2/2} du, \quad (5)$$

where  $C$  has mean value  $EC = m$  and variance  $\text{Var}C = s^2$ , and the logistic law

$$G(z) = \frac{1}{1 + e^{-z}}, \quad (6)$$

where  $C$  has mean value  $EC = m$  and variance  $\text{Var}C = \pi^2 s^2 / 3$ .

For both of these  $G(0) = F_C(m) = \Pr[y = 1 \mid v = m] = 1/2$ , and so  $m = V_{50}$ , the velocity for which the probability of penetration is 1/2. Note that  $m$  has two interpretations:

$$\begin{aligned} m &= EC \\ m &= V_{50}. \end{aligned} \quad (7)$$

The interpretation of  $s$  is the steepness of the ballistic response function, Eq. 3, proportional to the standard deviation of the critical velocity distribution, which relates to the size of a zmr. Based on the form of the normal or logistic probability laws,  $0 < p < 1$  for any  $v$ . So it is pointless to define the zmr as the velocity interval on which mixed results may be seen since this interval is  $(-\infty, \infty)$ , but the concept can be salvaged.

The inverse cdf or quantile function (qf) denoted  $Q = G^{-1}$  is used to express velocity as a function of penetration probability. Since  $Q(p) = (v - m)/s$ , one obtains  $V_{10}$  or  $V_{90}$  estimates, for example, by  $V_{100p} = m + s Q(p)$ .

So one can interpret the interval  $[V_{10}, V_{90}]$  as a sort of “80% zmr” in the sense that  $0.1 \leq p \leq 0.9$  in the interval,  $p < 0.1$  when  $v < V_{10}$ , and  $p > 0.9$  when  $v > V_{90}$ . The size of the interval is proportional to  $s$ . To complete the analogy, note that the classical zmr would be the open interval

$(V_0, V_{100})$ . For the example distributions, these quantities do not exist since  $\lim_{p \rightarrow 0^+} V_p = -\infty$  and  $\lim_{p \rightarrow 100^-} V_p = \infty$ .

Now  $s$  has 2 interpretations:

$$s \propto \sqrt{\text{Var } C}$$

$$s = \text{ballistic response steepness} = \text{zmr size}. \quad (8)$$

The parameter  $s$  characterizes the steepness of the response function or the standard deviation of the critical velocity distribution. It is not the “standard deviation of the ballistic limit”, it is not the “standard deviation of the  $V_{50}$ ”, and it is not used to provide confidence intervals or hypothesis tests on the  $V_{50}$ .

### 1.3 Properties of Maximum Likelihood Estimators

The population parameters  $m = V_{50}$  and  $s$  are fixed (but unknown) numbers to be estimated. They are not random. They do not have expected values, mean values, variances, standard deviations, or confidence intervals.

It is the estimates that are random. Analysis of this randomness enables quantification of error bounds associated with decisions based on such estimates. Decisions based on estimates devoid of error bounds cannot be justified.

Parameters are estimated from data  $(v_1, y_1), \dots, (v_n, y_n)$ , and the estimate  $\hat{m}$  of  $m$  is an estimate of the  $V_{50}$ . MLE provides one approach to point estimation, interval estimation, and inference based on the asymptotic (bivariate) normal distribution of parameter estimates. With  $\theta = (m, s)$  this is

$$\sqrt{n}(\hat{\theta} - \theta) \longrightarrow N_2(0, nM_\theta^{-1}) \quad \text{as } n \rightarrow \infty \quad (9)$$

where  $M_\theta$  is the information matrix. Then

$$\hat{\theta} \sim N_2(\theta, \text{Var } \hat{\theta}) \quad (10)$$

where  $\text{Var } \hat{\theta} = M_\theta^{-1}$ , and the asymptotic parameter variance structure is

$$\text{Var } \hat{\theta} = \begin{bmatrix} \text{Var } \hat{m} & \text{Cov}(\hat{m}, \hat{s}) \\ \text{Cov}(\hat{m}, \hat{s}) & \text{Var } \hat{s} \end{bmatrix}. \quad (11)$$

The asymptotic distribution of  $\hat{\theta}$  is normal for any critical velocity distribution: normal, logistic, or anything else. Confidence intervals and hypothesis tests on  $m = V_{50}$  are constructed with  $\hat{m}$  and  $\text{Var } \hat{m}$ , not  $\hat{s}$ .

This construction is possible only when the data has a zmr. Otherwise, none of the  $V_{50}$  estimate  $\hat{m}$ , its variance  $\text{Var } \hat{m}$ , the response steepness  $\hat{s}$ , the information matrix  $M_\theta$ , or anything else can be computed in the MLE framework.

When the data has a zmr, however, a reasonable and simple approach would involve a statistical hypothesis test on  $m (= V_{50})$  of the form

$$\begin{aligned} H_0 : m &= m_0 \\ H_1 : m &> m_0 \end{aligned} \tag{12}$$

with a specified Type I error level  $\alpha = 1 - \gamma$ . Rejection of the null hypothesis would lead to the conclusion that  $V_{50}$  meets or exceeds the specification value  $m_0$ . The probability that this decision is erroneous (concluding that  $V_{50} > m_0$  when in fact  $V_{50} < m_0$ ) would be no greater than  $\alpha$ . The risk of such a procedure is controlled by setting the value of  $\alpha$ . For a large enough data set, an accurate test can be based on the null asymptotic normal distribution of the estimator,  $\hat{m} \sim N(m, \text{Var } \hat{m})$ . In this case, the decision rule for the hypothesis test is

$$\text{reject } H_0 \iff \hat{m} > m_c \tag{13}$$

where the critical value for the test is  $m_c = m_0 + Z_\gamma \sqrt{\text{Var } \hat{m}}$ , and  $Z_\gamma$  is the appropriate standard normal quantile, e.g.,  $Z_{0.9} = 1.282$  for  $\alpha = 0.1$ , or  $Z_{0.95} = 1.645$  for  $\alpha = 0.05$ .

The naive decision rule of Eq. 1, equivalently,

$$\text{reject } H_0 \iff \hat{m} > m_0 \tag{14}$$

arises from  $Z_\gamma = 0$ , or  $\gamma = \alpha = 0.5$ , and in this case the probability of an erroneous rejection (concluding that  $V_{50} > m_0$  when in fact  $V_{50} < m_0$ ) is 50%.

The Generalized Linear Model (GLM) approach provides the necessary parameter and variance estimates. The linear representation of the simple logistic regression model is

$$p = \frac{1}{1 + e^{-a - bx}} \tag{15}$$

where the linear parameter vector is  $\beta = (a, b)$ . The usual location-scale parameterization is  $\theta = (m, s)$  where  $a + bx = (x - m)/s$ , so  $sa = -m$  and  $sb = 1$ . GLM computation provides the parameter estimate  $\hat{\beta} = (\hat{a}, \hat{b})$ , and thus  $\hat{\theta} = (\hat{m}, \hat{s})$  as indicated. For this data

x	691	700	701	715	717	718	719	720	729	742
y	0	0	0	1	1	0	0	0	1	1

the parameter estimates are

$$\hat{a} = -119.1315227 \quad \hat{b} = 0.1655234 \quad \hat{m} = 719.7262014 \quad \hat{s} = 6.0414421. \quad (16)$$

The information matrix

$$M_\beta = X^t W X = \begin{bmatrix} \sum w_i & \sum w_i x_i \\ \sum w_i x_i & \sum w_i x_i^2 \end{bmatrix} \quad (17)$$

is constructed from the design matrix  $X = [x_1^1 \dots x_n^1]^t$  and the diagonal weight matrix  $W = \text{diag}(w_1, \dots, w_n)$  where the weights are  $w_i = p_i(1 - p_i)$ . Then the asymptotic distribution of the linear parameter estimator is  $\hat{\beta} \sim N_2(\beta, \text{Var } \hat{\beta})$  where  $\text{Var } \hat{\beta} = M_\beta^{-1}$ . The location-scale parameter variance is obtained by using  $T = s [1 \ 0]$  in the reparameterization transformation  $\text{Var } \hat{\theta} = \hat{T}^t \text{Var } \hat{\beta} \hat{T}$ , and then the asymptotic distribution of the location-scale parameter estimator is  $\hat{\theta} \sim N_2(\theta, \text{Var } \hat{\theta})$ . The variances depend on the true (unknown) parameter values, but estimators are substituted for computational purposes.

### 1.3.1 Confidence Interval

An estimate  $\hat{M}_\beta$  of  $M_\beta$  is obtained using  $\hat{a}$  and  $\hat{b}$  to get the  $\hat{p}_i$ ,  $\hat{w}_i$ , and  $\hat{W}$ . Then  $\hat{\text{Var}} \hat{\beta} = \hat{M}_\beta^{-1}$ , and the estimated asymptotic distribution of  $\hat{\beta}$  is  $\hat{\beta} \sim N_2(\beta, \hat{\text{Var}} \hat{\beta})$ . The location-scale parameter variance estimate is obtained by using  $\hat{T} = \hat{s} [1 \ 0]$  in the reparameterization transformation  $\hat{\text{Var}} \hat{\theta} = \hat{T}^t \hat{\text{Var}} \hat{\beta} \hat{T}$ . This is enough to estimate

$$\hat{\sigma}^2 = \hat{\text{Var}} \hat{m} = \hat{s}^2 [1, \hat{m}] \hat{\text{Var}} \hat{\beta} [1, \hat{m}]^t = 26.25406, \quad (18)$$

which serves to construct a confidence interval (CI) on the unknown true parameter value  $m$ . Assuming  $\hat{m} \sim N(m, \hat{\sigma}^2)$ , then

$$\Pr[\hat{m} \in (-\infty, m + \hat{\sigma} Z_\gamma)] = \Pr[\hat{m} \leq m + \hat{\sigma} Z_\gamma] = \gamma. \quad (19)$$

Inverting this interval gives the equivalent  $\Pr[m \in I] = \gamma$  where the CI is  $I = [I_0, \infty)$  and  $I_0 = \hat{m} - \hat{\sigma}Z_\gamma$ . With  $\alpha = 0.1$ , so  $\gamma = 0.9$ , and  $Z_\gamma = 1.281552$ , the lower confidence bound is

$$I_0 = \hat{m} - \hat{\sigma}Z_\gamma = 713.1597. \quad (20)$$

### 1.3.2 CI-based Hypothesis Test

Consider the hypothesis test Eq. 12 with

$$m_0 = 712. \quad (21)$$

The usual approach is to apply the CI to the hypothesis test and reject if  $m_0$  lies outside the CI, which is  $m_0 < I_0$ , or equivalently if  $\hat{m} > m_c$  where critical value is

$$m_c = m_0 + \hat{\sigma}Z_\gamma = 718.5665. \quad (22)$$

In this case,  $m_0 < I_0$ , or  $\hat{m} > m_c$ , provide evidence to reject  $H_0$  at  $\alpha = 0.1$  and conclude that  $m > m_0$ . The p-value for the test is the  $\alpha$  corresponding to  $m_c = \hat{m}$ , which is

$$p = \Pr[N(m_0, \hat{\sigma}^2) > \hat{m}] = 0.06579217. \quad (23)$$

An equivalent decision rule is to reject  $H_0$  if  $p < \alpha$ , which is satisfied in this case.

### 1.3.3 Wald Test

With  $\theta = \begin{bmatrix} m \\ s \end{bmatrix}$ , the 2-sided test of

$$\begin{aligned} H_0 : \theta &= \theta_0 \\ H_1 : \theta &\neq \theta_0 \end{aligned} \quad (24)$$

is based on the Wald statistic

$$W = (\hat{\theta} - \theta_0)^t M_{\theta_0} (\hat{\theta} - \theta_0). \quad (25)$$

Since  $\hat{\theta} \sim N_2(\theta_0, V_{\theta_0})$  with  $V_{\theta_0} = M_{\theta_0}^{-1}$  under  $H_0$ , the null distribution of  $W$  is chi-squared

$$W = (\hat{\theta} - \theta_0)^t M_{\theta_0} (\hat{\theta} - \theta_0) \sim \chi_2^2 \quad (26)$$

and MLE provides the variance estimate  $M_{\theta_0} = \hat{M}_{\theta}$ . This is the Cramér-Rao (CR) variance bound, and the nomenclatures Wald and CR are interchangeable. Under  $H_1$  with  $\theta = \theta_1$  and  $\hat{\theta} \sim N_2(\theta_1, V_{\theta_1})$ , note that  $\hat{\theta} - \theta_0 = \hat{\theta} - \theta_1 + (\theta_1 - \theta_0)$ . Then, the distribution of  $W$  is noncentral chi-squared with noncentrality parameter  $\delta = (\theta_1 - \theta_0)^t M_{\theta_1} (\theta_1 - \theta_0)$ , that is to say

$$W = (\hat{\theta} - \theta_0)^t M_{\theta_1} (\hat{\theta} - \theta_0) \sim \chi^2_{2,\delta}. \quad (27)$$

For any linear transformation  $K$ , the test

$$\begin{aligned} H_0 : K(\theta - \theta_0) &= 0 \\ H_1 : K(\theta - \theta_0) &\neq 0 \end{aligned} \quad (28)$$

is based on the null normal distribution of  $K\hat{\theta} \sim N_k(K\theta_0, KV_{\theta_0}K^t)$  where  $\text{rank } K = k$  and the null chi-squared distribution of  $W$ , namely

$$W = (K(\hat{\theta} - \theta_0))^t (KV_{\theta_0}K^t)^{-1} (K(\hat{\theta} - \theta_0)) \sim \chi^2_k. \quad (29)$$

Under  $H_1$  with  $\theta = \theta_1$  and  $\hat{\theta} \sim N_2(\theta_1, V_{\theta_1})$ , the distribution of  $W$  is noncentral chi-squared with noncentrality parameter  $\delta = (K(\theta_1 - \theta_0))^t (KV_{\theta_1}K^t)^{-1} (K(\theta_1 - \theta_0))$ , which is

$$W = (K(\hat{\theta} - \theta_0))^t (KV_{\theta_1}K^t)^{-1} (K(\hat{\theta} - \theta_0)) \sim \chi^2_{k,\delta}. \quad (30)$$

To test

$$\begin{aligned} H_0 : m &= m_0 \\ H_1 : m &\neq m_0 \end{aligned} \quad (31)$$

take  $K = [1, 0]$ , which implies that  $k = 1$ . Then, under  $H_0$

$$W = \frac{(\hat{m} - m_0)^2}{\text{Var}_0 \hat{m}} \sim \chi^2_1 \quad (32)$$

with  $\text{Var}_0 \hat{m} = \hat{\text{Var}} \hat{m}$ . The critical value for the test is the appropriate chi-squared quantile

$$W_c = Q_{\chi^2_1}(\gamma), \quad (33)$$

and the p-value for the test is

$$p = \Pr[\chi_1^2 > W] = 1 - F_{\chi_1^2}(W). \quad (34)$$

Under  $H_1$ , the distribution of  $W$  is noncentral chi-squared with noncentrality parameter  $\delta(m_1) = (m_1 - m_0)^2 / \text{Var}_1 \hat{m}$ ,

$$W = \frac{(\hat{m} - m_0)^2}{\text{Var}_1 \hat{m}} \sim \chi_{1,\delta(m_1)}^2. \quad (35)$$

Type II error  $\beta(m_1)$  is the probability of erroneously failing to reject  $H_0$ . This depends on a specific alternative parameter value  $m = m_1$ . The power of the test is  $\pi(m_1) = 1 - \beta(m_1)$ , the probability of correctly rejecting  $H_0$ . Then, the power of this test is

$$\pi(m_1) = \Pr[\chi_{1,\delta(m_1)}^2 > W_c] = 1 - F_{\chi_{1,\delta(m_1)}^2}(W_c). \quad (36)$$

The Wald statistic for the example data is

$$W = \frac{(\hat{m} - m_0)^2}{\hat{\text{Var}} \hat{m}} = \frac{(719.7262014 - 712)^2}{26.25406} = 2.273713 \quad (37)$$

where  $\hat{\text{Var}} \hat{m}$  estimates  $\text{Var}_0 \hat{m}$ . Since the Wald test is naturally 2-sided, divide by 2 to get its 1-sided p-value, which is

$$\frac{\Pr[\chi_1^2 > W]}{2} = 0.06579217 \quad (38)$$

as expected. The 2-sided error rate is  $\alpha_2 = 2\alpha = 0.2$ , and then  $\gamma_2 = 1 - \alpha_2 = 0.8$ , which means the Wald statistic critical value is

$$W_c = Q_{\chi_1^2}(\gamma_2) = 1.642374, \quad (39)$$

and  $H_0$  is rejected with  $\hat{m} > m_0$  when  $W > W_c$  or  $p < \alpha$ . Since the square of a normal is chi-squared, the Wald and CI-based tests are equivalent. One-sided Wald tests can be conducted if the error levels are adjusted and the direction of variation is considered. The CI-based test is just another way of looking at the Wald test.

A small-sample simulation illustrates aberrant behavior of the Wald/CI logistic regression test power. See Fig. 1. This is a structural property of the Wald test and is not caused by “bad data” or “small samples”. The asymptotic power computation (Eq. 36) exhibits this behavior but not to such an extent as the simulated small-sample power. Sample size is  $n = 11$  and evenly-spaced values of  $x$  on  $[-10, 10]$  are used throughout. True null parameters are  $m = 0$  and  $s = 2$ .

Two-sided tests (Eq. 31) with  $m_0 = 0$  are conducted with  $\alpha = 0.2$ . The null and 50 values of  $m_1$  evenly spaced on  $[-10, 10]$  are used and  $s = 2$  throughout, so alternatives are shifts of the true null. For each parameter value, 10,000 tests are simulated. Power is computed as the relative frequency of rejections for each  $m$ . Power should increase,  $\pi \rightarrow 1$ , as  $|m_1 - m_0|$  increases, but the power of the Wald test actually decreases for large  $|m_1 - m_0|$  and eventually  $\pi \rightarrow \alpha$ . This type of behavior was reported as early as 1977 by Hauck & Donner<sup>10</sup> for the linear parameterization. Tests based on the likelihood ratio (LR), to be developed later in this report, behave correctly even for this small sample size.

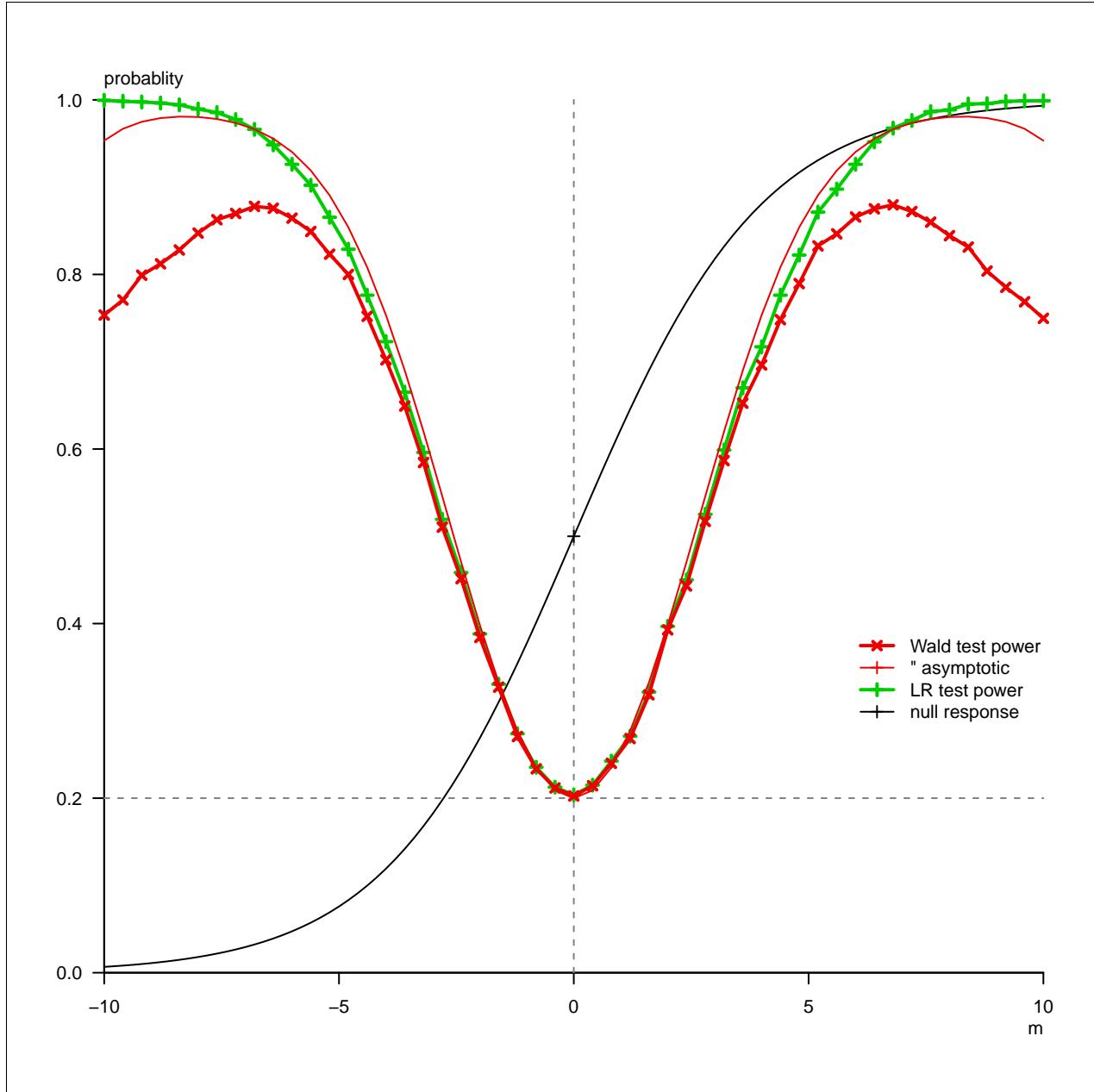


Fig. 1 Small-sample and asymptotic test power

### 1.3.4 Wald (CR) CIs on Mean Response

The usual construction of CIs on mean response  $E[y|x]$  is based on the linear parameterization  $P(x) = G(x\beta)$  where  $\beta^t = [b_0, b_1]$ . We can write  $x = [1, x]$  without confusion. The asymptotic parameter distribution is  $\hat{\beta} \sim N_2(\beta, V)$  with  $V = V_\beta = (X^t W X)^{-1}$  as in section 1.3. Since

$x\hat{\beta} \sim N(x\beta, xVx^t)$  we take the CI to be

$$G\left(x\hat{\beta} \pm \sqrt{xVx^t} \cdot Z\right) \quad (40)$$

for some normal quantile  $Z$ . For large enough  $\gamma$ , these CIs increase in size for extreme  $x$  as shown in Fig. 2, where the red curves are CR CIs and the blue curves are LR CIs. We see that CR CIs always have this undesirable property. The equation for the CI band is

$$\frac{(y - x\hat{\beta})^2}{xVx^t} = Q_{\chi_1^2}(\gamma) \quad (41)$$

so we write

$$(y - x\hat{\beta})^2 - xTx^t = 0 \quad (42)$$

where  $\beta^t = [A, B]$  and  $T = Q_{\chi_1^2}(\gamma) \cdot V = \begin{bmatrix} C & D \\ D & E \end{bmatrix}$  and then recognize the equation of an hyperbola

$$\begin{aligned} 0 &= (y - A - Bx)^2 - C - 2Dx - Ex^2 \\ &= (B^2 - E)x^2 - 2Bxy + y^2 + 2(AB - D)x - 2Ay + (A^2 - C) \\ &= ax^2 + bxy + cy^2 + dx + ey + f. \end{aligned} \quad (43)$$

The gradient vanishes at the center  $(x_o, y_o)$ , thus  $(2ax_o + by_o + d, 2cy_o + bx_o + e) = (0, 0)$ . So  $(x_o, y_o) = (2cd - be, 2ae - bd)/(b^2 - 4ac)$ . Here,  $b^2 - 4ac = 4B^2 - 4(B^2 - E) = 4E$  and  $2cd - be = 4(AB - D) - 4AB = -4D$  and  $2ae - bd = 4A(E - B^2) + 4B(AB - D) = 4(AE - BD)$ . Since  $V = (X^tWX)^{-1}$  and  $W$  is diagonal, the center is

$$\begin{bmatrix} x_o \\ y_o \end{bmatrix} = \begin{bmatrix} -D/E \\ A - BD/E \end{bmatrix} = \begin{bmatrix} \Sigma wx/\Sigma w \\ A + B\Sigma wx/\Sigma w \end{bmatrix}. \quad (44)$$

Hyperbola asymptotes pass through the center. Asymptote slopes come from  $ax^2 + bxy + cy^2 = 0$ , or  $y = (-bx \pm \sqrt{b^2x^2 - 4acx^2})/(2c) = x(-b \pm \sqrt{b^2 - 4ac})/(2c)$ , so the slopes are

$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2c} = B \pm \sqrt{E} = b_1 \pm \sqrt{Q_{\chi_1^2}(\gamma) \cdot V_{2,2}} \quad (45)$$

where  $V_{2,2} = \Sigma w / [\Sigma w \Sigma wx^2 - (\Sigma wx)^2]$ . We can always choose  $\gamma$  large enough to make an asymptote slope negative and elicit the undesirable behavior by taking  $\gamma > F_{\chi_1^2}(b_1^2/V_{2,2})$ .

The elements of  $W$  are  $w = G(x\beta)[1 - G(x\beta)]$ , so this can be computed from the regression coefficients. Figure 3 shows the linearized version of Fig. 2.

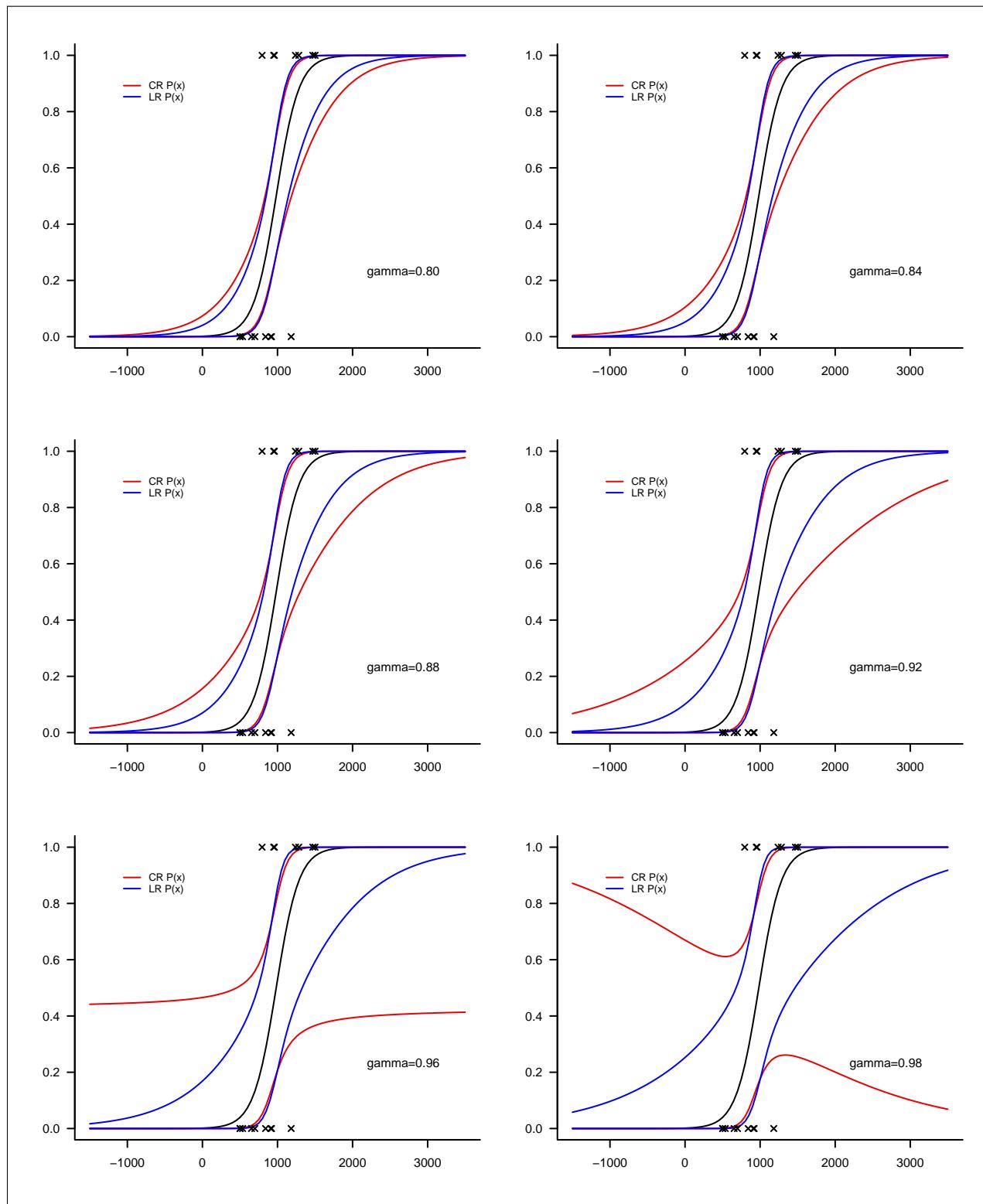


Fig. 2 CR and LR CIs, various  $\gamma$

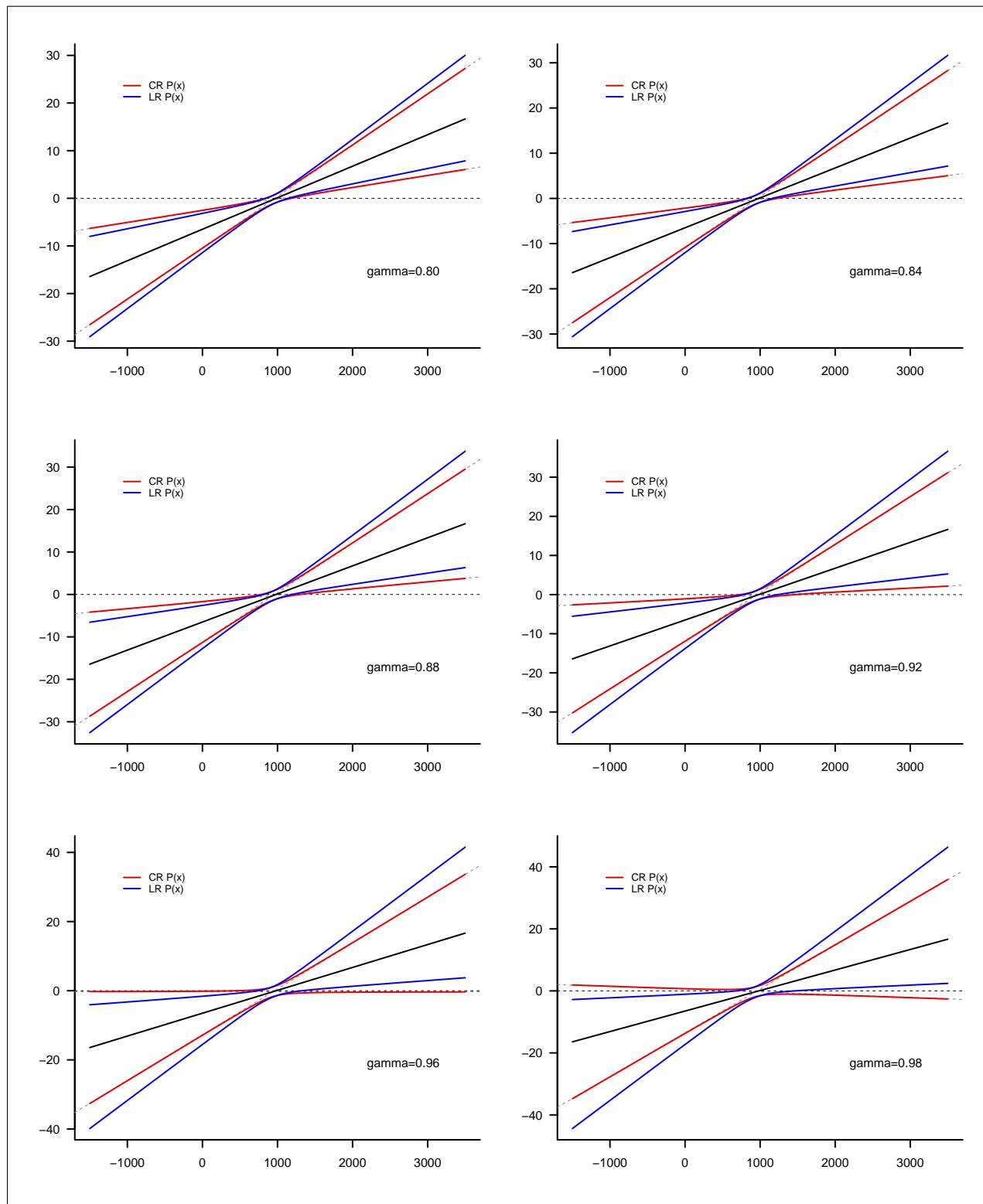


Fig. 3 Linearized CR and LR CIs, various  $\gamma$

## 1.4 Suggested Approach

The computations needed to do the correct test in the asymptotic MLE framework were provided by Golub and Grubbs<sup>1</sup> in 1950. A detailed modern approach including portable estimation and generalized inference computation based on the asymptotic MLE methodology is described by Collins.<sup>11</sup>

The MLE approach has limitations. Confidence intervals can be distorted, test power is not consistent, and the method does not apply at all to data with no zmr.

At least 2 authors already provide approaches for analyzing data with no zmr (gap data).

Webb<sup>12</sup> uses random noise data augmentation to produce zmr data (and therefore parameter point estimates) from data with no zmr. He uses simulation to estimate parameter errors. He makes no claim that these computed quantities actually represent the true population parameters of interest.

In fact, hypothesis testing is based on interval estimates. Point estimation is unnecessary, and statistically valid acceptance testing can therefore be based conceptually on interval estimation without recourse to point estimation.

Neyer,<sup>13</sup> a self-proclaimed “provider of Efficient Sensitivity Test and Analysis Software”, produces closed proprietary software that implements LR methodology, thus allowing inference for data with no zmr. Since this software is closed and the theoretical documentation incomplete, it is impossible to verify the details of his work.

The LR approach provides the needed technology. The LR approach to quantal response model inference allows analysis of data with or without a zmr on equal ground. LR test power is consistent, unlike the Wald test. The asymptotic distribution of the LR test statistic is easy to compute. It is the aim of this report to present the relevant LR theory and methodology in an open and complete manner, and thus encourage discussion, criticism, and implementation by the larger community.

---

## 2. Likelihood Ratio Estimation and Inference

---

### 2.1 Likelihood

Consider a random sample of  $n$  independent and identically distributed (iid) data  $z_1, \dots, z_n$  from a distribution with fixed unknown parameter  $\theta \in \Theta$  where  $\Theta$  is the space of possible parameter values. The likelihood function is the joint probability density function (pdf) of the sample, considered as a function of the parameter

$$L(\theta) = \prod_{i=1}^n f(z_i; \theta). \quad (46)$$

A related quantity is the deviance  $D$ , used in theoretical development and software diagnostics,

$$D = -2 \log L. \quad (47)$$

So  $L$  can be expressed as

$$L = e^{-D/2}, \quad (48)$$

knowledge of either determines the other, and

$$D(\theta) = -2 \sum_{i=1}^n \log f(z_i; \theta). \quad (49)$$

An MLE  $\hat{\theta}$  of  $\theta$ , if it exists, satisfies

$$L(\hat{\theta}) = \sup\{L(\theta) : \theta \in \Theta\} = \sup_{\theta \in \Theta} L(\theta) = \sup_{\Theta} L, \quad (50)$$

which is the same as

$$D(\hat{\theta}) = \inf\{D(\theta) : \theta \in \Theta\} = \inf_{\theta \in \Theta} D(\theta) = \inf_{\Theta} D, \quad (51)$$

so maximum likelihood is equivalent to minimum deviance.

### 2.2 Likelihood Ratio Tests

Stuart<sup>14</sup> provides details on various tests using the likelihood ratio. Simple hypotheses (null or alternative) completely specify the distribution.

For  $\theta_0, \theta_1 \in \Theta$ , a uniformly most powerful test for the simple hypothesis against the simple alternative

$$\begin{aligned} H_0 : \theta &= \theta_0 \\ H_1 : \theta &= \theta_1 \end{aligned} \tag{52}$$

is based on the simple likelihood ratio (SLR)

$$\Lambda = \frac{L(\theta_0)}{L(\theta_1)}. \tag{53}$$

For  $\Theta_o \subseteq \Theta$ , a test of the possibly composite hypothesis

$$\begin{aligned} H_0 : \theta &\in \Theta_o \\ H_1 : \theta &\notin \Theta_o \end{aligned} \tag{54}$$

can be based on the generalized likelihood ratio (GLR)

$$\Lambda = \frac{\sup_{\theta \in \Theta_o} L(\theta)}{\sup_{\theta \in \Theta} L(\theta)}. \tag{55}$$

Consider  $\Theta = \{\theta_0, \theta_1\}$  and  $\Theta_o = \{\theta_0\}$ , in which case  $\theta \in \Theta_o$  means  $\theta = \theta_0$  and  $\theta \notin \Theta_o$  means  $\theta = \theta_1$ . In this case the hypotheses of Eq. 54 are both simple, but the GLR

$$\Lambda = \frac{L(\theta_o)}{\sup\{L(\theta_o), L(\theta_1)\}} \tag{56}$$

does not coincide with the SLR of Eq. 53. However, the GLR test can be formulated in situations where the SLR test cannot, and the GLR test is good enough to be useful. We use the GLR test exclusively. GLR, Eq. 55, tests are equivalent to tests based on the deviation

$$\Delta = -2 \log \Lambda \tag{57}$$

so

$$\Delta = \inf_{\theta \in \Theta_o} D(\theta) - \inf_{\theta \in \Theta} D(\theta). \tag{58}$$

Note that

$$\Lambda = e^{-\Delta/2}. \tag{59}$$

## 2.3 Statistic Distribution

The probability distribution of the test statistic  $\Lambda$  is required for hypothesis tests, Type I error, p-values, and CIs.

### 2.3.1 Asymptotics

Wilks<sup>15</sup> gives the large-sample distribution of  $\Delta$

$$\Delta \sim \chi_r^2 \quad (60)$$

as chi-squared with  $r = \dim \Theta - \dim \Theta_o$  degrees of freedom.

## 2.4 Hypothesis Test Construction

The critical value of  $\Lambda$  for a test of Eq. 54 with Type I error  $\alpha = 1 - \gamma$  is

$$\Lambda^* = Q_\Lambda(\alpha) \quad (61)$$

since small  $\Lambda$ , Eq. 55, indicates significant departure from  $H_0$ . The decision is based on the observed (estimated) value  $\hat{\Lambda}$  of the likelihood ratio

$$\text{reject } H_0 \text{ if } \hat{\Lambda} < \Lambda^*. \quad (62)$$

The p-value  $\hat{p}$  for the test is the probability that  $\Lambda$  is as extreme as the observed value  $\hat{\Lambda}$ , so

$$\hat{p} = F_\Lambda(\hat{\Lambda}). \quad (63)$$

An equivalent decision rule is

$$\text{reject } H_0 \text{ if } \hat{p} < \alpha, \quad (64)$$

since

$$\hat{\Lambda} < \Lambda^* \iff \hat{\Lambda} < Q_\Lambda(\alpha) \iff F_\Lambda(\hat{\Lambda}) < F_\Lambda(Q_\Lambda(\alpha)) \iff \hat{p} < \alpha. \quad (65)$$

In terms of the deviation  $\Delta = -2 \log \Lambda$ , large  $\Delta$ , Eq. 58, is significant. The critical value is

$$\Delta^* = Q_\Delta(\gamma), \quad (66)$$

the decision is

$$\text{reject } H_0 \text{ if } \hat{\Delta} > \Delta^*, \quad (67)$$

and the p-value is

$$\hat{p} = 1 - F_{\Delta}(\hat{\Delta}). \quad (68)$$

Using the large-sample asymptotic approximation, Eq. 60, the critical value of  $\Delta$  is

$$\Delta^* = Q_{\chi_r^2}(\gamma), \quad (69)$$

and the p-value is

$$\hat{p} = 1 - F_{\chi_r^2}(\hat{\Delta}). \quad (70)$$

## 2.5 Single Point

With  $\Theta_o = \{c\}$ , we can test that the parameter has a fixed value

$$\begin{aligned} H_0 : \theta &= c \\ H_1 : \theta &\neq c. \end{aligned} \quad (71)$$

Suppose the test is constructed with Type I error  $\alpha = 1 - \gamma$ . The critical value  $\Lambda^*(\alpha)$  depends on  $\alpha$  but not  $c$  or the data. The test statistic  $\hat{\Lambda}(c)$  depends on  $c$  and the data but not  $\alpha$ , and likewise for the p-value

$$\hat{p}(c) = F_{\Lambda}(\hat{\Lambda}(c)). \quad (72)$$

### 2.5.1 Confidence Region

The set  $C$  of  $c$  for which the test fails to reject

$$C = \{c : \hat{\Lambda}(c) \geq \Lambda^*(\alpha)\} = \{c : \hat{p}(c) \geq \alpha\} \quad (73)$$

serves as a  $\gamma$  confidence region for  $\theta$ .

### 2.5.2 Confidence Interval

If  $\theta$  is a scalar, then there are 2 values  $c_0$  and  $c_1$  of  $c$  with

$$c_0 < \hat{\psi} < c_1 \quad (74)$$

where  $\hat{p}(c) \geq \alpha$  for  $c_0 \leq c \leq c_1$ , and  $\hat{p}(c) < \alpha$  for  $c < c_0$  or  $c > c_1$ , so

$$\hat{p}(c_0) = \hat{p}(c_1) = \alpha, \quad (75)$$

and the  $\gamma$  confidence region for  $\theta$  is an interval

$$C = [c_0, c_1]. \quad (76)$$

Note that  $\hat{p}(\hat{\theta}) = \hat{\Lambda}(\hat{\theta}) = 1$ .

---

### 3. Application to Quantal Response

---

#### 3.1 The Quantal Response Model

In the quantal response (QR) model,<sup>11</sup> the response is a binary (Bernoulli) random variable

$$y \in \{0, 1\} \quad (77)$$

with Bernoulli parameter  $p = P(x) = \Pr[y = 1 | x] = E[y | x]$  dependent on a stimulus  $x \in \mathbb{R}$

$$P(x) = G(a + bx) \quad (78)$$

through a function  $G$  called the link and an unknown linear model parameter  $\theta = (a, b)$ . Let  $Q$  denote the inverse of  $G$ . Then the stimulus  $x = X(p)$  with a given mean response  $p$  is the inverse of  $P$ . With  $p = G(a + bx)$ , then  $Q(p) = a + bx$  and

$$X(p) = \frac{Q(p) - a}{b}. \quad (79)$$

Common choices for the link  $G$  include the logistic cdf for the logit model (logistic regression)

$$G(z) = \frac{1}{1 + e^{-z}} \quad (80)$$

with inverse

$$Q(u) = \log \left( \frac{u}{1 - u} \right) \quad (81)$$

and the normal cdf function for the probit model (probit regression)

$$G(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-u^2/2} du \quad (82)$$

with inverse

$$Q(u) = \inf\{z : G(z) \geq u\}. \quad (83)$$

Useful parameterizations of the model are

$$P(x) = G(b(x - m)) = G\left(\frac{x - m}{s}\right). \quad (84)$$

where  $bs = 1$  and  $a = -bm$ . Then

$$X(p) = m + \frac{Q(p)}{b} = m + sQ(p). \quad (85)$$

If  $G(0) = 1/2$ , as it does for the logit and probit models, then  $P(m) = 1/2$  also, and so  $m$  is the stimulus level at which the probability of response equals  $1/2$ , that is,  $m = X(1/2)$ .

In ballistic work where  $x$  is velocity and  $y$  is penetration,  $m = X(1/2)$  is the  $V_{50}$ , or that velocity with a 50% chance of penetration. In other words,  $P(V_{50}) = 1/2$ .

## 3.2 Estimation

Suppose we have a data set  $\{(x_1, y_1), \dots, (x_n, y_n)\}$  of size  $n$ . Let  $x_{0\max} = \max\{x_i : y_i = 0\}$ , the highest stimulus with no response. Let  $x_{1\min} = \min\{x_i : y_i = 1\}$ , the lowest stimulus with a response. If  $x_{1\min} < x_{0\max}$ , the data set has a zmr  $Z = [x_{1\min}, x_{0\max}]$ . If  $x_{0\max} \leq x_{1\min}$ , the data set has a (possibly empty) gap  $Z = (x_{0\max}, x_{1\min})$  with no data.

### 3.2.1 With a Zone of Mixed Results

Data with a zmr admit a unique MLE with a smooth S-shaped response curve

$$P(x) = G(a + bx) = G(b(x - m)) = G\left(\frac{x - m}{s}\right) \quad (86)$$

and likelihood  $0 < L < 1$ , or deviance  $D > 0$  (Fig. 4).

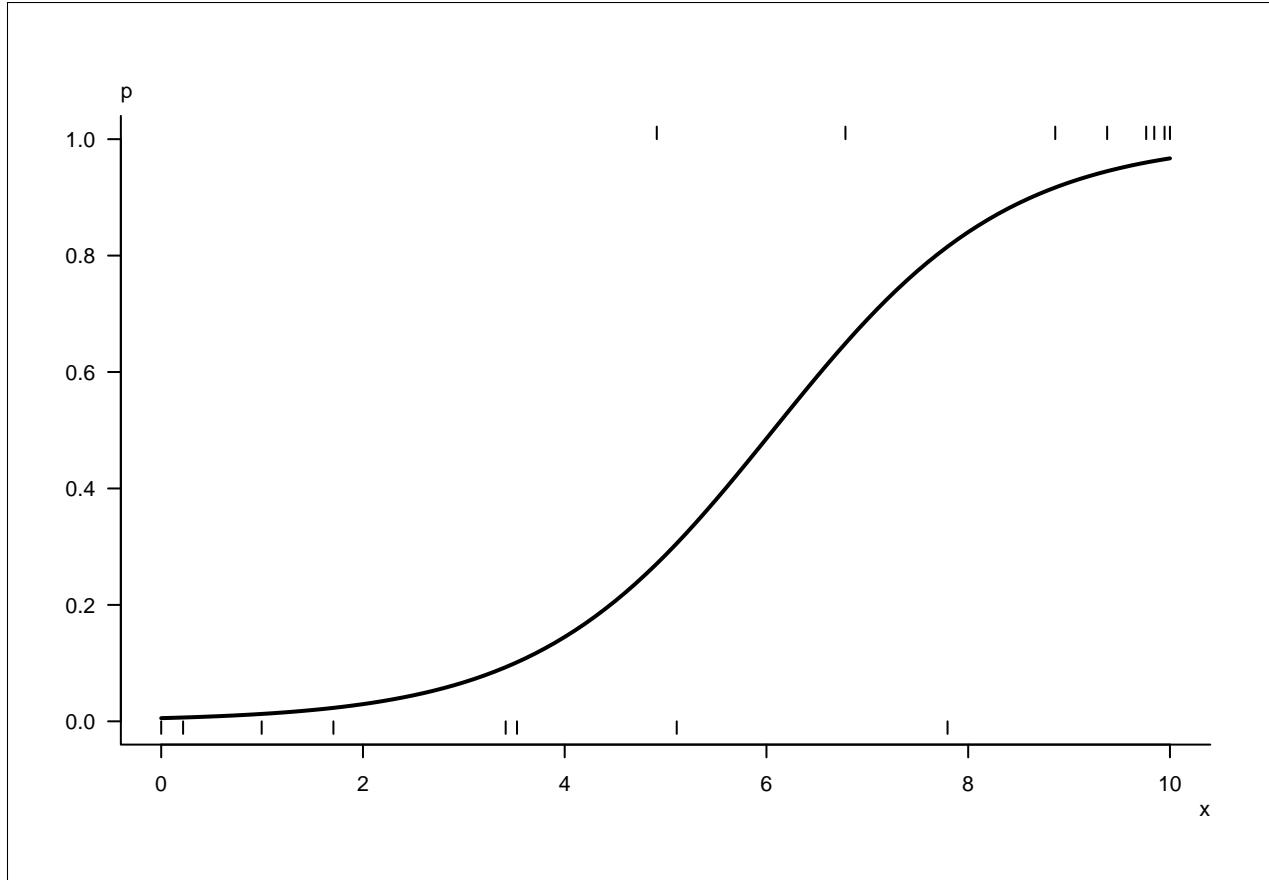


Fig. 4 Zmr data

Estimation of  $\theta$  in the QR model is best implemented using GLM. The computation is detailed in Collins<sup>11</sup> for implementation in any programming language.

For general or development work, it is convenient to use an interactive environment that already implements the necessary computations, such as R,<sup>16</sup> S,<sup>17</sup> S-PLUS, Mathematica,<sup>18</sup> MATLAB,<sup>19</sup> or JMP.<sup>20</sup> In R (or S or S-PLUS), the stimulus and response are in a data frame

```
z <- data.frame(x=x, y=y)
```

and the estimation for Eqs. 78 and 80

```
fit <- glm(y~x, family=binomial(link="logit"), data=z)
```

returns an object containing  $\theta$  in

```
fit$coefficients
```

and the deviance  $D$  as

```
D <- fit$deviance .
```

### 3.2.2 With No Zone of Mixed Results

Note that  $y = 0$  at  $x = x_{0\max}$ , and  $y = 0$  if  $x < x_{0\max}$ .

$$P(x) = \begin{cases} 0, & x < x_{0\max} \\ 1, & x > x_{1\min} . \end{cases} \quad (87)$$

Figure 5 illustrates gap data.

In terms of the specific model with parameterizations of Eqs. 78 and 84, one usually considers a fixed  $m \in Z$  and a sequence of  $b_i$  with  $b_i \rightarrow \infty$  as  $i \rightarrow \infty$ . Then the sequences of parameters  $(a, b) = (-mb_i, b_i)$  for Eq. 78 or  $(m, s) = (m, 1/b_i)$  for Eq. 84 are equivalent, and they have  $L_i \rightarrow 1$  and  $D_i \rightarrow 0$  also. The limiting 2-parameter model is the step function (indeterminate at  $x = m$ )

$$P(x) = \begin{cases} 0, & x < m \\ ?, & x = m \\ 1, & x > m \end{cases} \quad (88)$$

or simply

$$P(x) = \begin{cases} 0, & x < m \\ 1, & x > m \end{cases} \quad (89)$$

with  $L = 1$  and  $D = 0$ . Even though the MLE does not exist and the limiting form is not unique, these values of  $L$  and  $D$  can be used in computation, estimation, and inference.

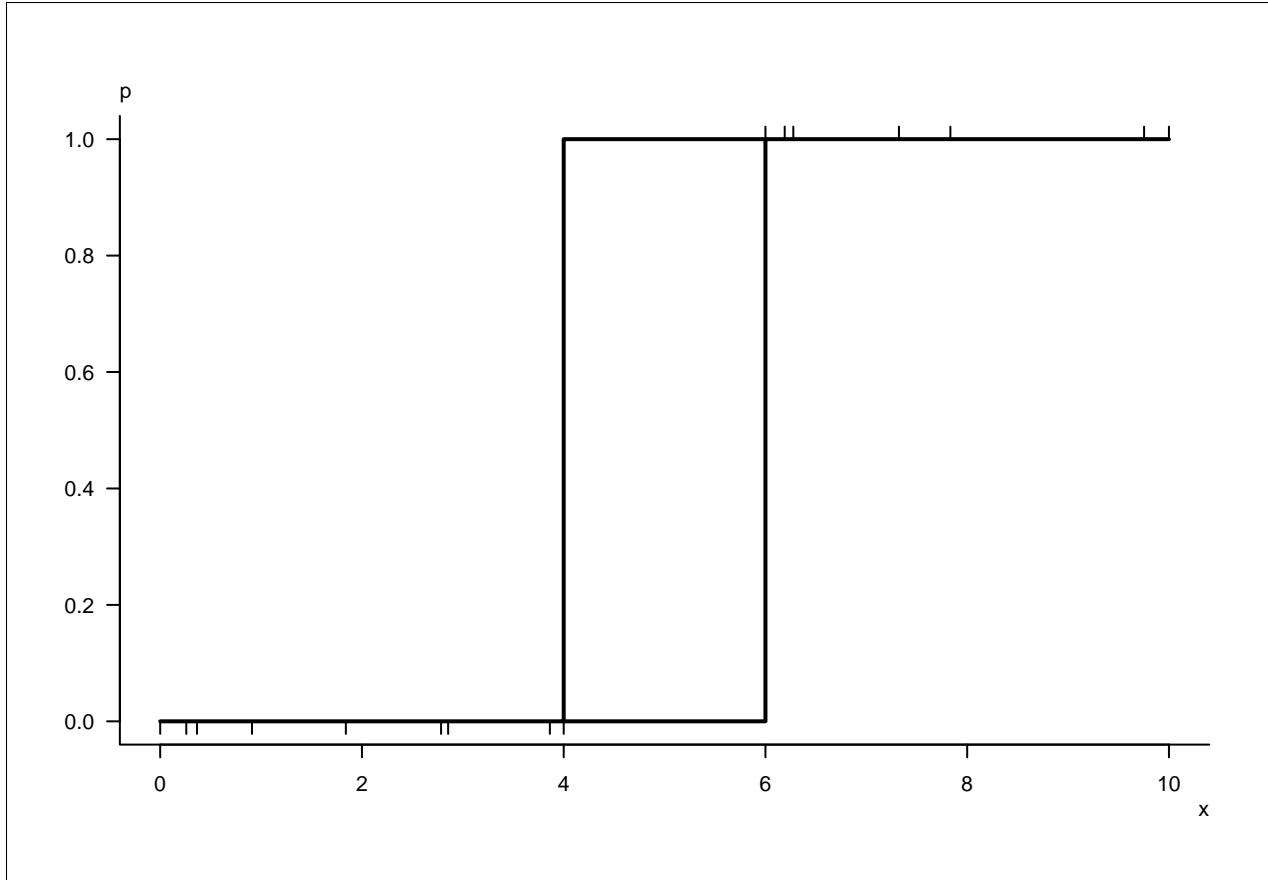


Fig. 5 Gap data

### 3.3 One-Sample Inference

In this section, all tests are 2-sided GLR tests, so  $H_1$  is the negation of  $H_0$ .

One special case is

$$H_0 : a = a_o \quad (90)$$

in the model  $P(x) = G(a + bx)$ . Since  $G(0) = a$ , this is the same as

$$H_0 : P(0) = a_o . \quad (91)$$

Another special case is

$$H_0 : X(0.5) = m_o . \quad (92)$$

In ballistic terminology,  $X(0.5) = V_{50}$ . This even works with no zmr. If the data set has a zmr,

then the model can be written as  $P(x) = G((x - m)/s)$  and this test can be written as

$$H_0 : m = m_o. \quad (93)$$

In any case, this test is the same as

$$H_0 : P(m_o) = 0.5. \quad (94)$$

In general, the fundamental 1-sample test for the QR model is that the response curve passes through a certain point  $(x_o, p_o)$

$$H_0 : P(x_o) = p_o \quad (95)$$

or, equivalently,

$$H_0 : X(p_o) = x_o. \quad (96)$$

We can construct the corresponding GLR test of Eq. 71, evaluate the deviation  $\Delta = D_o - D$ , Eq. 58, and obtain the p-value  $\hat{p}$ , Eq. 68.

The denominator term  $D$  (from  $\theta \in \Theta$ ) is the deviance of the unrestricted model MLE. The numerator term  $D_o$  (from  $\theta \in \Theta_o$ ) is the deviance of the restricted model MLE.

If the data has no zmr and  $x_o$  is in the gap, then  $L_o = L = 1$  and  $D_o = D = 0$ , so  $\Lambda = 1$  and  $\Delta = 0$  and  $\hat{p} = 1$ . Otherwise, the data has a zmr, or the data has no zmr and  $x_o$  is not in the gap. To estimate the restricted numerator model and obtain  $D_o$ , consider the parameterization

$$P(x) = G(q_o + b(x - x_o)) \quad (97)$$

where

$$q_o = Q(p_o) \quad (98)$$

and note that

$$P(x_o) = G(q_o + b(x_o - x_o)) = G(q_o + b \cdot 0) = G(q_o) = p_o. \quad (99)$$

Estimation of the single parameter  $b$  in this model, restricting the response to pass through  $(x_o, p_o)$ , produces the required  $L_o$  and  $D_o$ .

This yields inference and interval estimation on QR parameters even in the case of no zmr.

If the data set has no zmr, then technically  $L = 1$  and  $D = 0$ . This gives confidence intervals which may appear “too narrow”. Furthermore, the p-value is not a continuous function of the data in the sense that an infinitesimal shift in one  $x_i$  can change the data set between zmr and gap and

produce a (non-infinitesimal) jump in p-value. Neyer<sup>13</sup> suggests the following correction that makes the p-value continuous and gives larger (more conservative) confidence intervals.

### 3.3.1 Continuity Correction

Consider that the maximum value of the likelihood function is in fact  $L = 1/4$  when the data has a zmr, and so the minimum deviance is  $D = \log 16$ . To see this, the simplest example is a sample of size  $n = 2$  with  $x_1 < x_2$  and  $y_1 = 1$  and  $y_2 = 0$ . With respect to an increasing response  $G(x_1) \leq G(x_2)$ , this data has a zmr. Let  $z_i = G(x_i)$ . The likelihood function is

$$L = \prod_{i=1}^n G(x_i)^{y_i} (1 - G(x_i))^{1-y_i} = z_1(1 - z_2) = z_1 - z_1 z_2 \quad (100)$$

and we find its maximum value subject to the constraints  $0 \leq z_1 \leq z_2 \leq 1$ . The Kuhn-Tucker necessary conditions for the solution of  $\max L$  subject to  $g_j \leq 0$  are  $\nabla(L - \sum m_j g_j) = 0$  and  $m_j \geq 0$  and  $m_j g_j \geq 0$ . In this case,  $g_1 = -z_1$  and  $g_2 = z_1 - z_2$  and  $g_3 = z_2 - 1$ . Since  $L(0, \cdot) = L(\cdot, 1) = 0$  and  $L(1/2, 1/2) = 1/4$ , which is in fact the solution, we eliminate  $z_1 = 0$  and  $z_2 = 1$  and must take  $m_1 = m_3 = 0$ . Now  $m_2 = 0$  implies  $\nabla L = (1 - z_2, -z_1) = (0, 0)$  which is impossible, so  $m_2 = m > 0$  and  $z_1 = z_2 = z$ . Then  $(1 - z, -z) - m(1, -1) = (0, 0)$ , so  $z = m$  and  $1 - 2z = 0$  and  $z = 1/2$  as required. Any data set with a zmr contains 2 such points, and any larger data set only introduces additional factors in  $L$  which reduce its maximum value, since  $0 \leq G \leq 1$ .

Using these values  $L = 1/4$  and  $D = \log 16$  instead of  $L = 1$  and  $D = 0$  for the denominator with no zmr and  $x_o$  not in the gap makes the p-value of Eqs. 95 or 96 a continuous function of the data.

The resulting CIs are more conservative (larger, wider), as a more extreme numerator likelihood is necessary to obtain the required ratio. This correction is only applied to gap data (never to data with a zmr).

### 3.3.2 Implementation

To implement the computation in R, first transform (shift) the data

$$\tilde{x} = x - x_o \quad (101)$$

so  $\tilde{x}_o = 0$ . Then the restricted model

$$P(\tilde{x}) = G(q_o + b\tilde{x}) = G(q_o + b(x - x_o)) \quad (102)$$

has a fixed intercept  $a = q_o$  and a single parameter  $b$  to estimate (the slope).

To implement the computation in R and obtain the deviance  $D_o$ ,

```
zo <- data.frame(x=x-xo, y=y)
fito <- glm(y~x -1 +offset(qo), family=binomial(link="logit"), data=zo)
Do <- fito$deviance .
```

The new data frame  $zo$  holds the transformed stimulus. The  $-1$  term in the formula removes the intercept calculation, forcing the regression through  $(0,0)$ . The `offset` term imposes an offset, ultimately forcing the regression through  $(0, q_o)$ . Then the deviation  $\Delta$  is

```
Delta <- Do - D.
```

This may be inefficient and numerically unstable particularly when exhaustive calculation is required. Explicit solution is straightforward. See section 5.3.

To apply the large-sample  $\chi^2_r$  approximation for the distribution of  $\Delta$ , necessary only when the data set has a zmr or the data set has no zmr and  $x_o$  is not in the gap, one needs the degrees of freedom  $r$ . In the full (denominator) model,

$$\Theta = \{(a, b) : a \in \mathbb{R}, b \in \mathbb{R}\} \quad (103)$$

and  $\dim \Theta = 2$ , whether or not the data set has a zmr. In the restricted (numerator) model, since

$$a + bx = q_o + b(x - x_o) = (q_o + bx_o) + bx \quad (104)$$

we have

$$\Theta_o = \{(q_o + bx_o, b) : b \in \mathbb{R}\} \quad (105)$$

and  $\dim \Theta_o = 1$ . There is one free parameter  $b$ , and the value of  $b$  determines the value of  $a$ . For the shifted model

$$\tilde{\Theta}_o = \{(q_o, b) : b \in \mathbb{R}\}, \quad (106)$$

and  $\dim \tilde{\Theta}_o = 1$  also. Note that shifting does not affect the deviance

$$\tilde{D}_o = D_o. \quad (107)$$

Thus, the degrees of freedom for the asymptotic distribution of the deviation is  $r = 1$ , and the Wilks p-value is

$$\hat{p} = 1 - F_{\chi^2_r}(\Delta), \quad (108)$$

implemented in R as

```
p <- 1 - pchisq(Delta, 1).
```

Figures 6 and 7 illustrate examples of the zmr test and gap test, respectively.

Kinsler and Collins<sup>21</sup> provide examples of 1-sample tests.

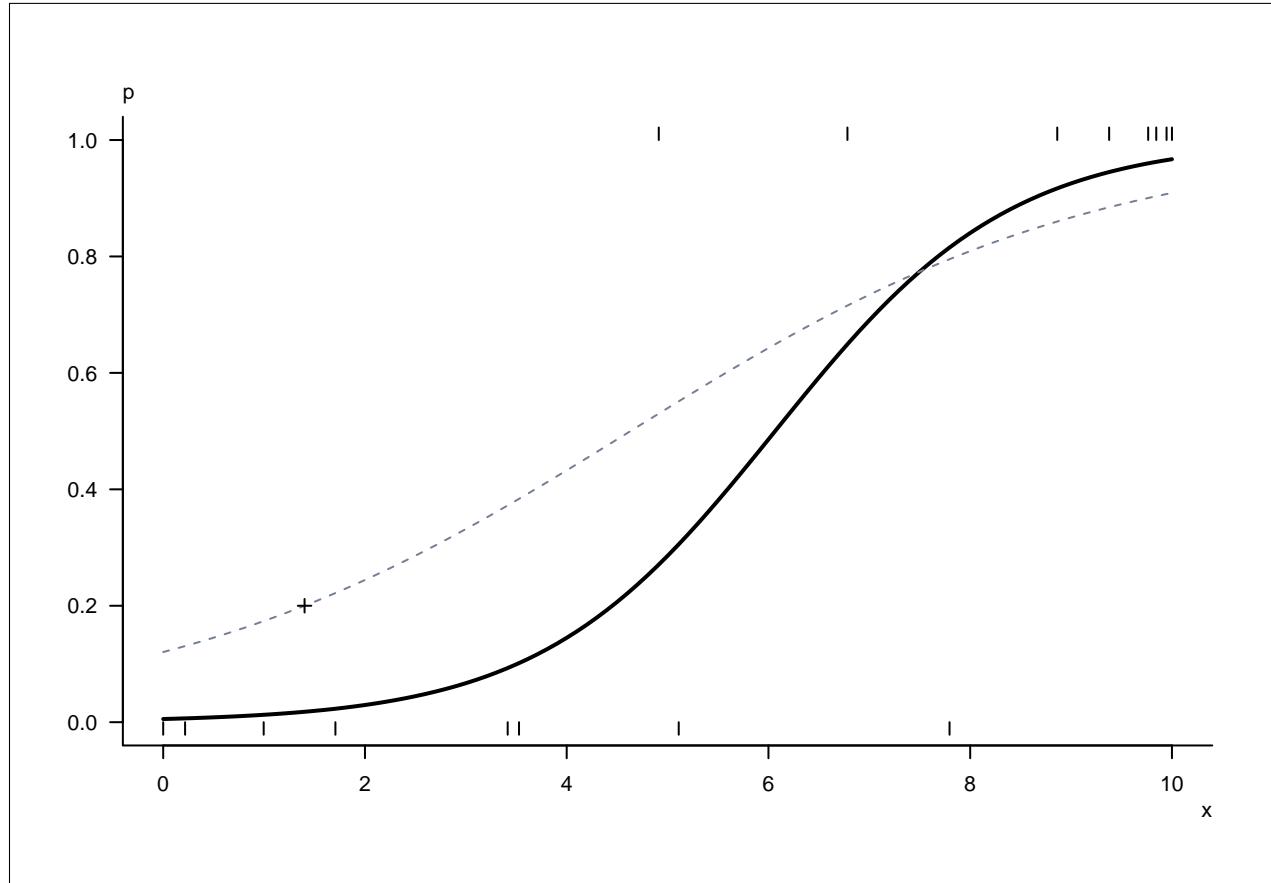


Fig. 6 Example of  $P(x_o) = p_o$  with zmr data

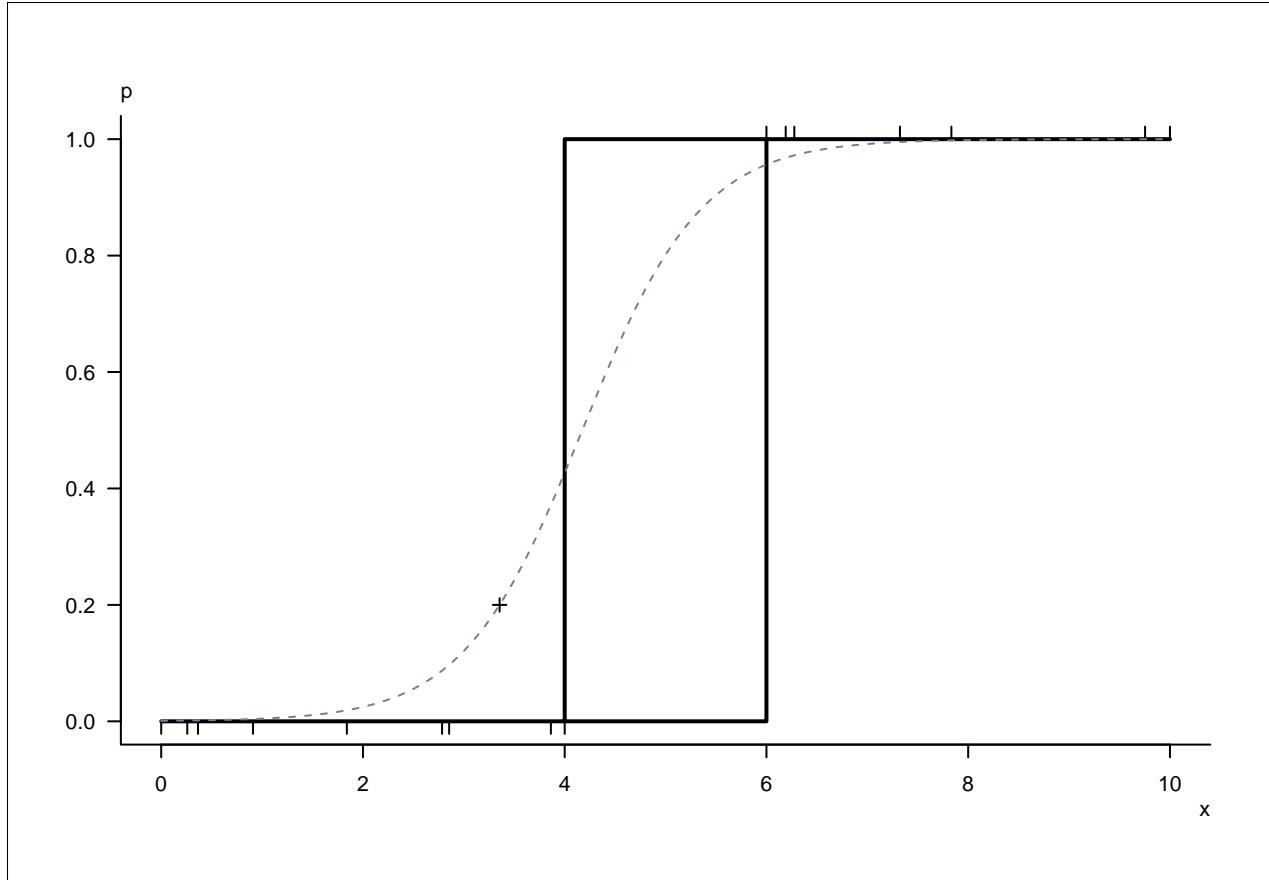


Fig. 7 Example of  $P(x_o) = p_o$  with gap data

### 3.3.3 Likelihood Ratio Confidence Intervals

In the QR model

$$P(x) = G(a + bx), \quad (109)$$

2-sided LR confidence bands for fixed  $x_o$  or  $p_o$  are obtained from the hypothesis test that the response curve passes through a certain point  $(x_o, p_o)$

$$\begin{aligned} H_0 &: P(x_o) = p_o \\ H_1 &: P(x_o) \neq p_o \end{aligned} \quad (110)$$

or, equivalently,

$$\begin{aligned} H_0 : X(p_o) &= x_o \\ H_1 : X(p_o) &\neq x_o. \end{aligned} \quad (111)$$

The restricted model uses the parameterization

$$P(x) = G(q_o + b(x - x_o)) \quad (112)$$

where

$$q_o = Q(p_o) \quad (113)$$

since

$$P(x_o) = G(q_o + b(x_o - x_o)) = G(q_o + b \cdot 0) = G(q_o) = p_o. \quad (114)$$

Estimation of the single parameter  $b$  in this model, restricting the response to pass through  $(x_o, p_o)$ , produces the likelihood  $L_o$ . The full model (usual estimation with 2 unrestricted parameters) yields likelihood  $L$ . The 2-sided p-value  $p$  for the test comes from the asymptotic distribution of the likelihood ratio  $\Lambda = L_o/L$

$$-2 \log \Lambda \sim \chi^2_1. \quad (115)$$

The 1-sided p-value is  $p/2$ .

In the ballistic application with  $p_o = 1/2$ , this is a test that  $V_{50} = x_o$ .

CIs are obtained by fixing  $p_o$  or  $x_o$  and adjusting the other to obtain the desired p-value.

### 3.3.4 Confidence Interval Examples

The following examples of LR and CR CIs on  $P(x)$  and  $X(p)$  for zmr and gap data illustrate the preceding discussion.

- Application of Eq. 76 to Eq. 95 yields confidence intervals on the probability of response  $p$  for given stimulus  $x$ . Figure 8 shows the multiple 1-sample tests used in the construction of CIs. Figures 9 and 10 show LR CIs ( $\gamma = 0.9$ ) on  $P$  for zmr and gap data, respectively.
- Application of Eq. 76 to Eq. 96 yields LR CIs on the stimulus  $x$  for given probability of response  $p$ . Figures 11 and 12 show LR CIs ( $\gamma = 0.9$ ) on  $X$  for zmr and gap data, respectively.
- Equivalence of LR CIs ( $\gamma = 0.9$ ) on  $P$  and  $X$  is illustrated in Figs. 13 and 14 for zmr and gap data, respectively.
- LR CIs for multiple  $\gamma$ , (0.8, 0.9, 0.95, 0.99) are illustrated in Figs. 15–18. Figure 15 shows CIs for zmr data. Figures 16 and 17 show CIs for gap data without and with the continuity correction, respectively. Figure 18 shows CIs for gap data both with and without the continuity correction to illustrate its effect.
- Comparisons of LR to CR CIs ( $\gamma = 0.9$ ) on  $P$  and  $X$  for zmr data are illustrated in Figs. 19–22. Figures 19 and 20 show CR CIs on  $P$  and  $X$ , respectively. Figure 21 shows CR CIs on both  $P$  and  $X$ , while Fig. 22 shows CR and LR CIs on  $P$  and  $X$ .
- CR CIs show the effect explained in section 1.3.4. Figures 23–25 use data provided by David W Webb, Mathematical Statistician, US Army Research Laboratory, Weapons and Materials Research Directorate, Advanced Weapons Concepts Branch, with  $\gamma = 0.95$ . Figures 23 and 24 show LR and CR CIs, respectively. Figure 25 shows LR and CR CIs together. LR CIs can exhibit similar behavior. See Fig. 26 for examples with  $\gamma = 0.99$  and random samples of size 16. In general, the effect is less likely and less extreme for LR CIs.

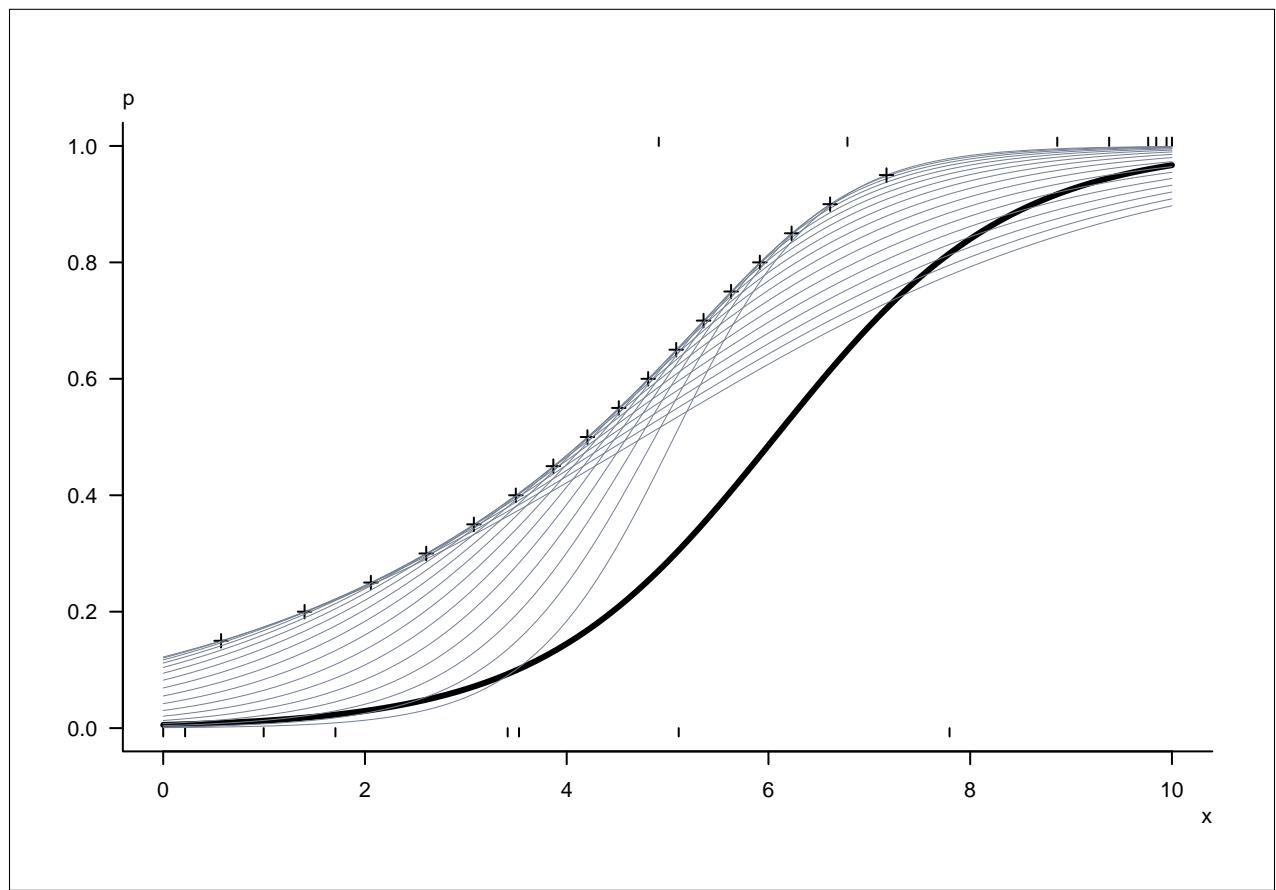


Fig. 8 Multiple  $P(x_o) = p_o$  with zmr

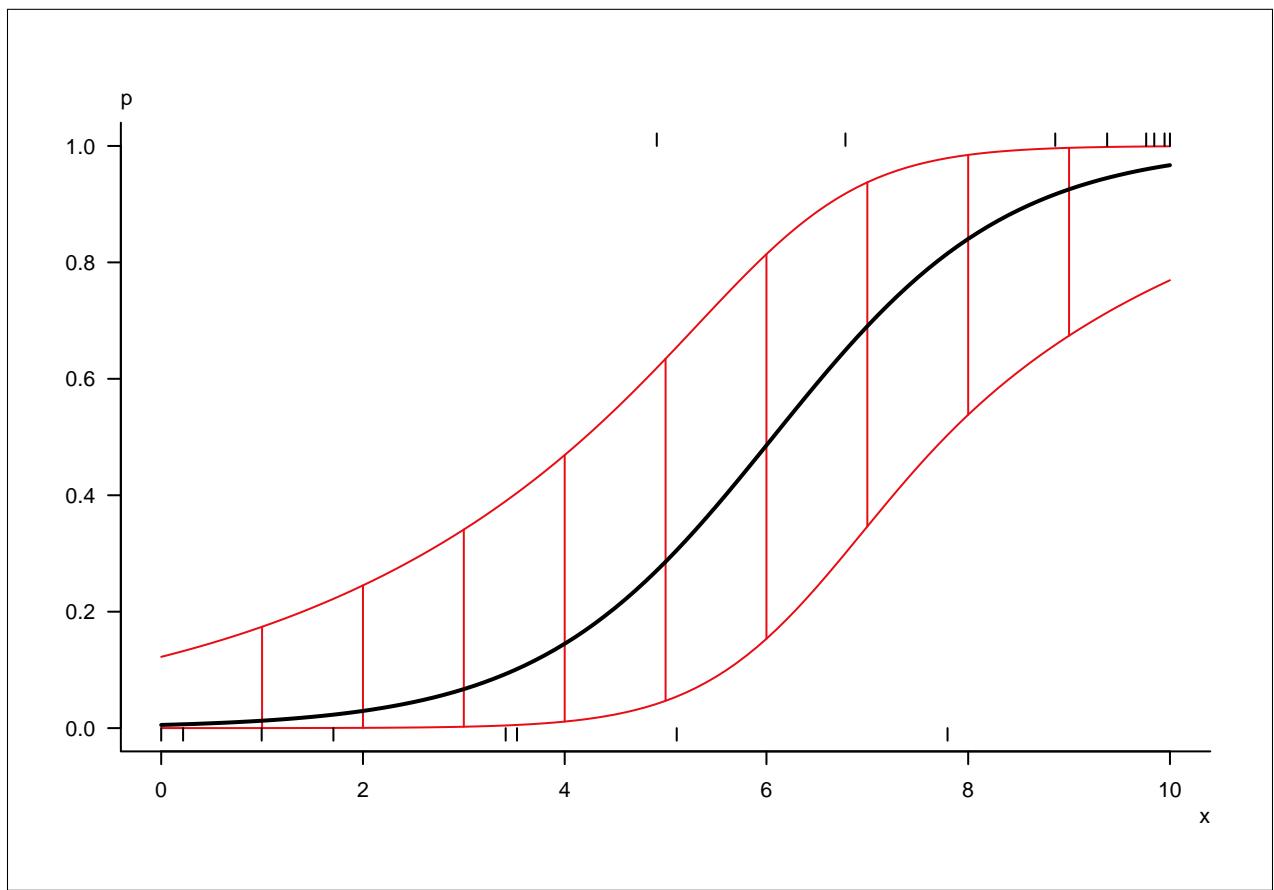


Fig. 9 LR CIs on  $P(x)$  for zmr data

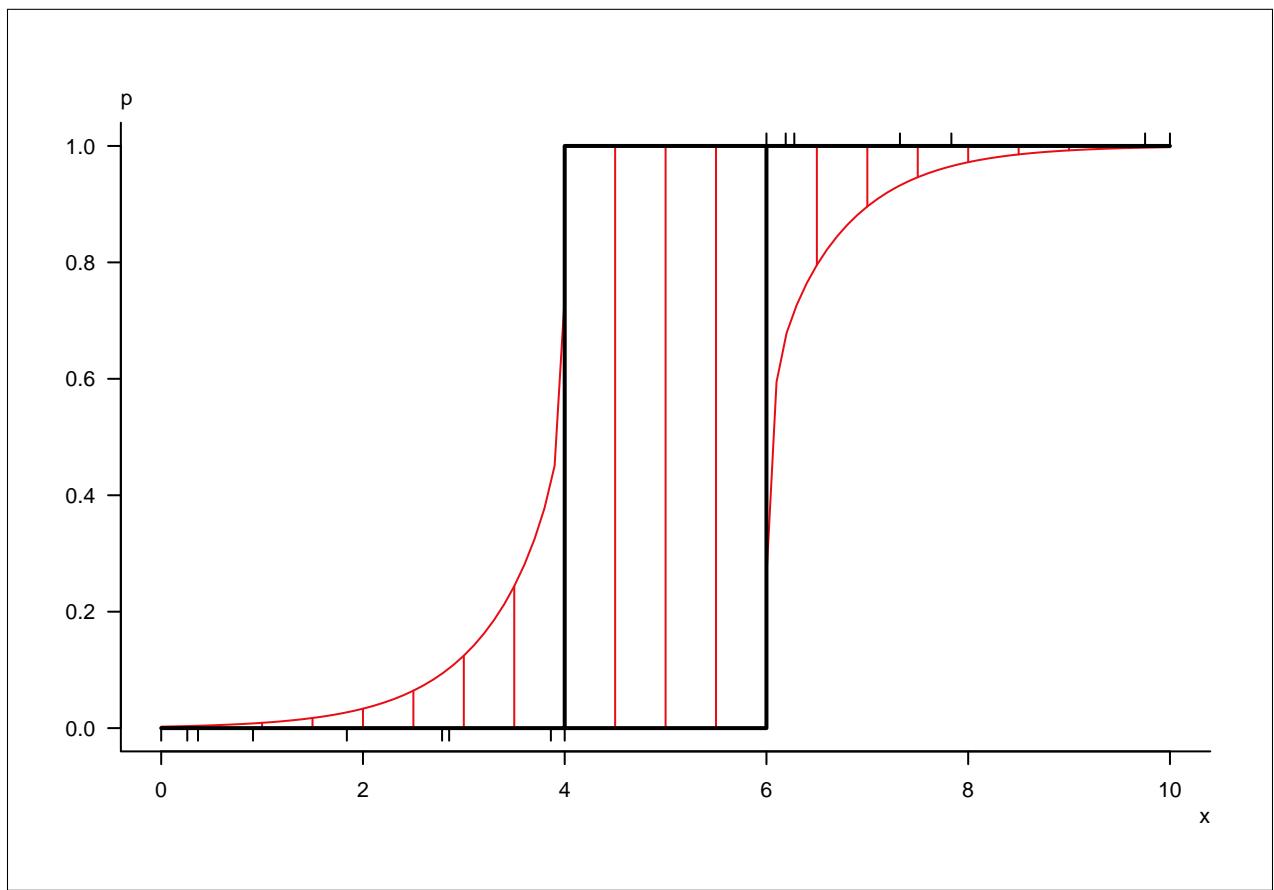


Fig. 10 LR CIs on  $P(x)$  for gap data

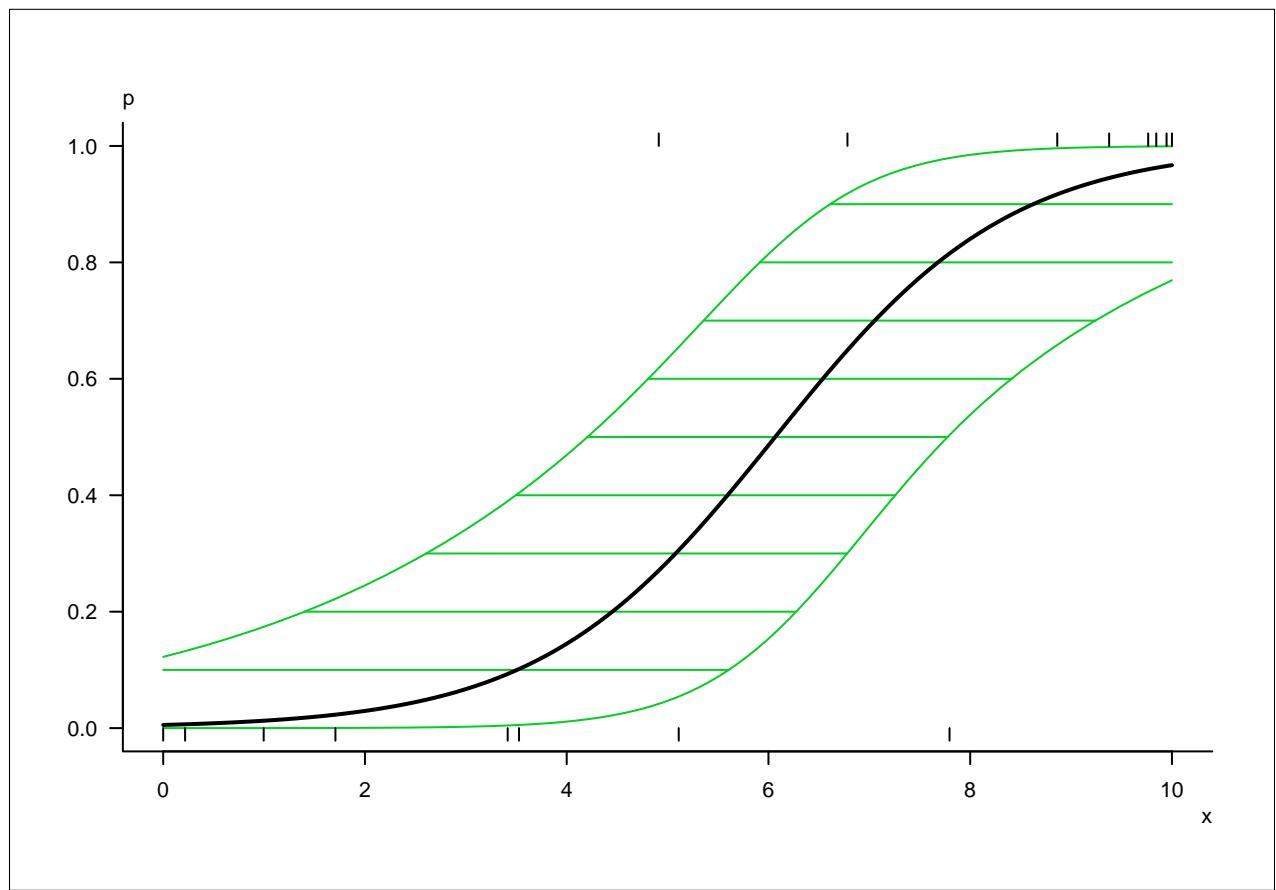


Fig. 11 LR CIs on  $X(p)$  for zmr data

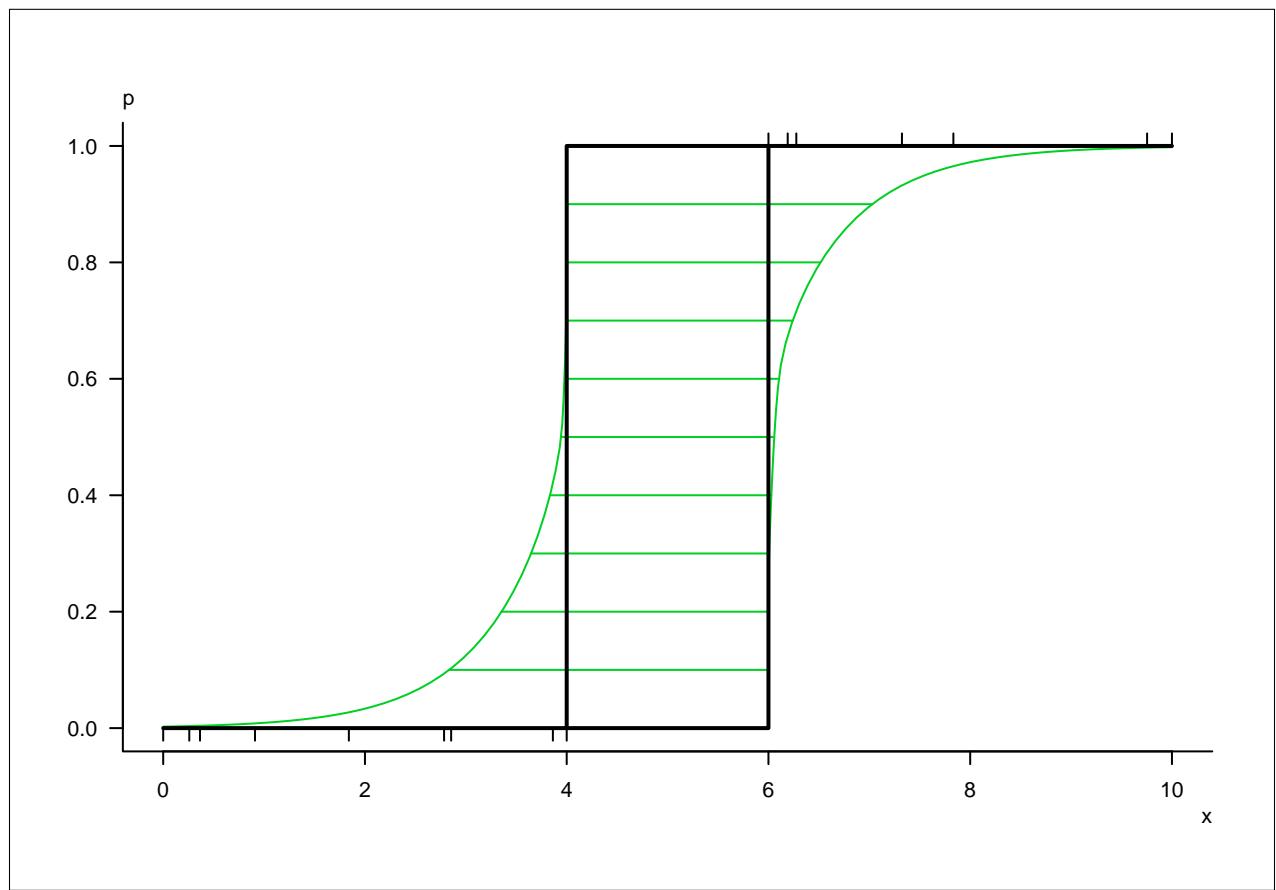


Fig. 12 LR CIs on  $X(p)$  for gap data

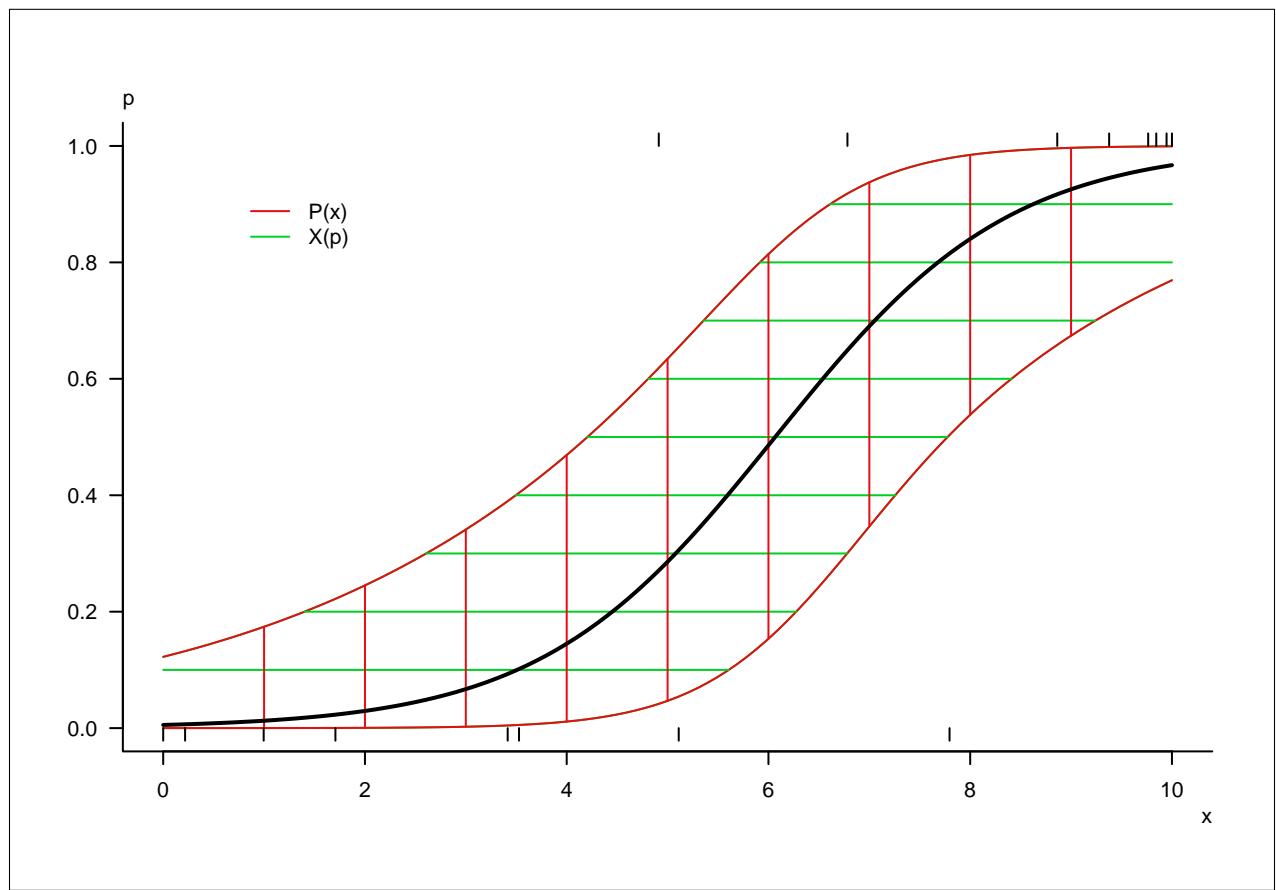


Fig. 13 LR CIs on  $P(x)$  and  $X(p)$  for zmr data

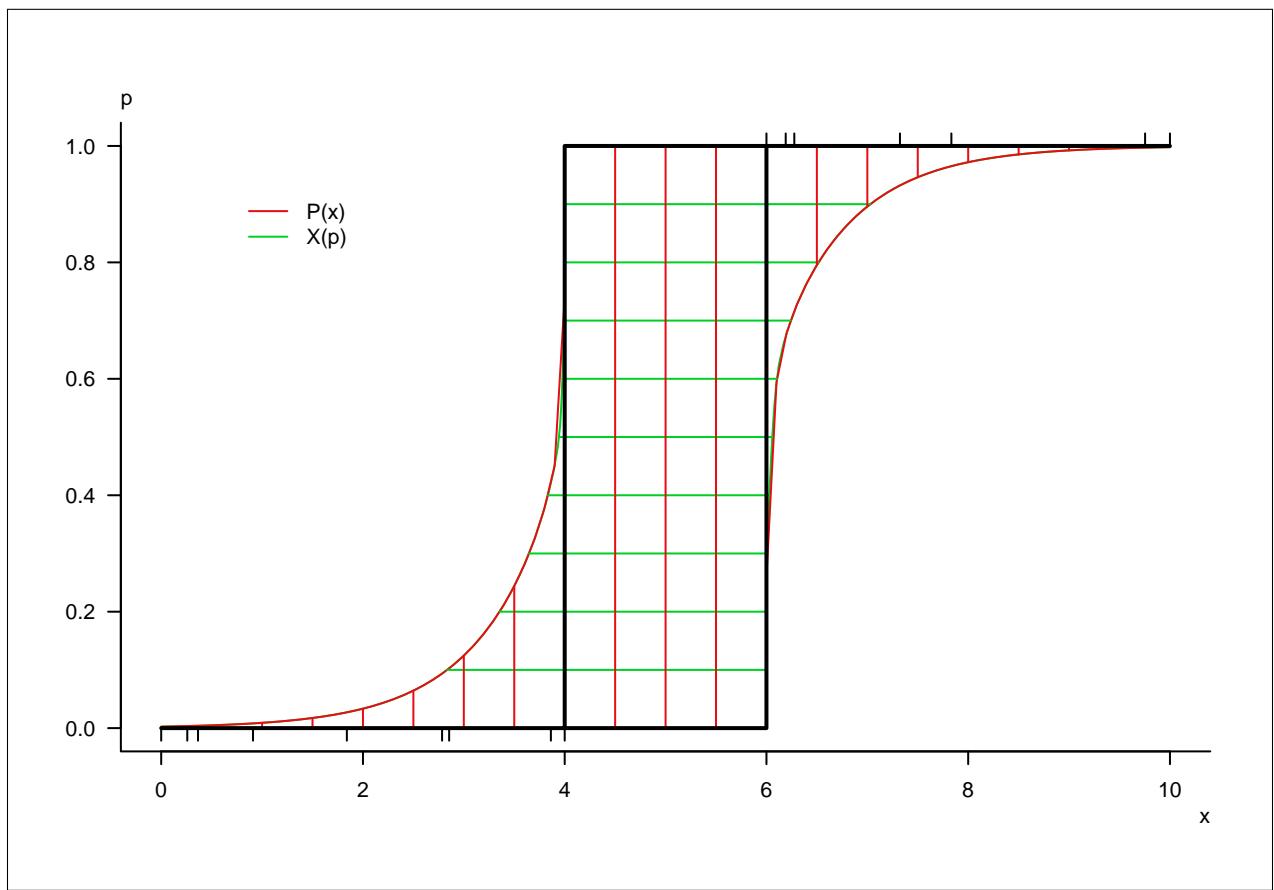


Fig. 14 LR CIs on  $P(x)$  and  $X(p)$  for gap data

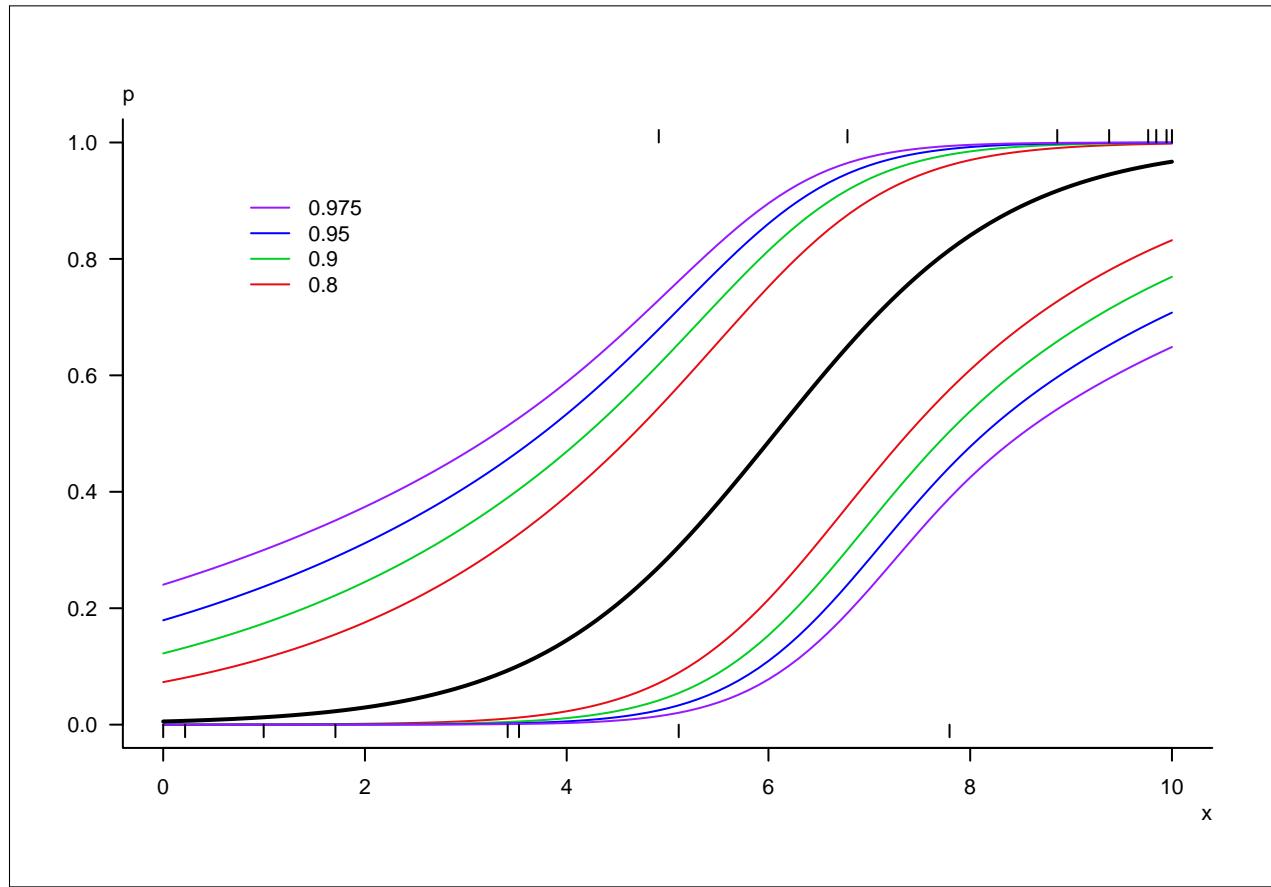


Fig. 15 LR CIs bands for zmr data

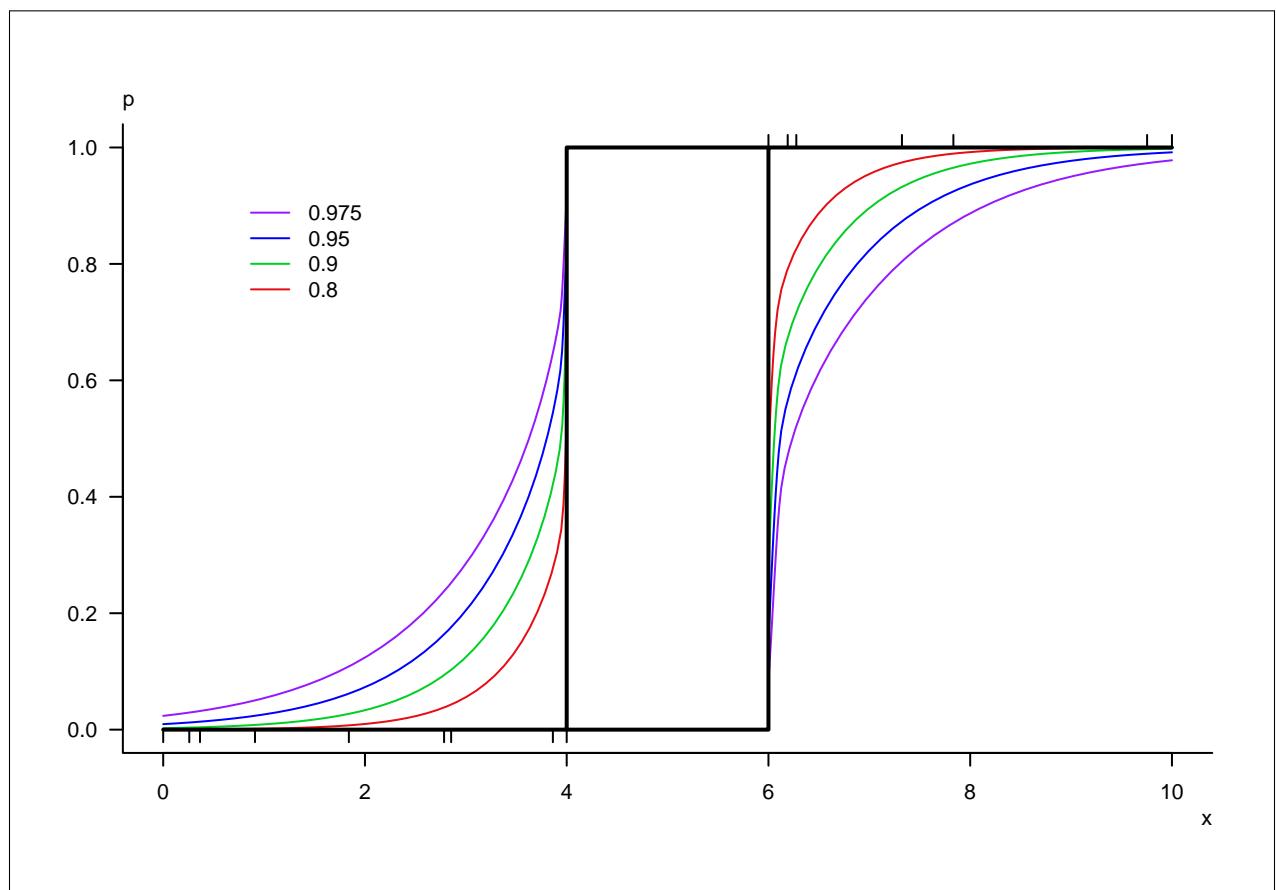


Fig. 16 LR CI bands for gap data

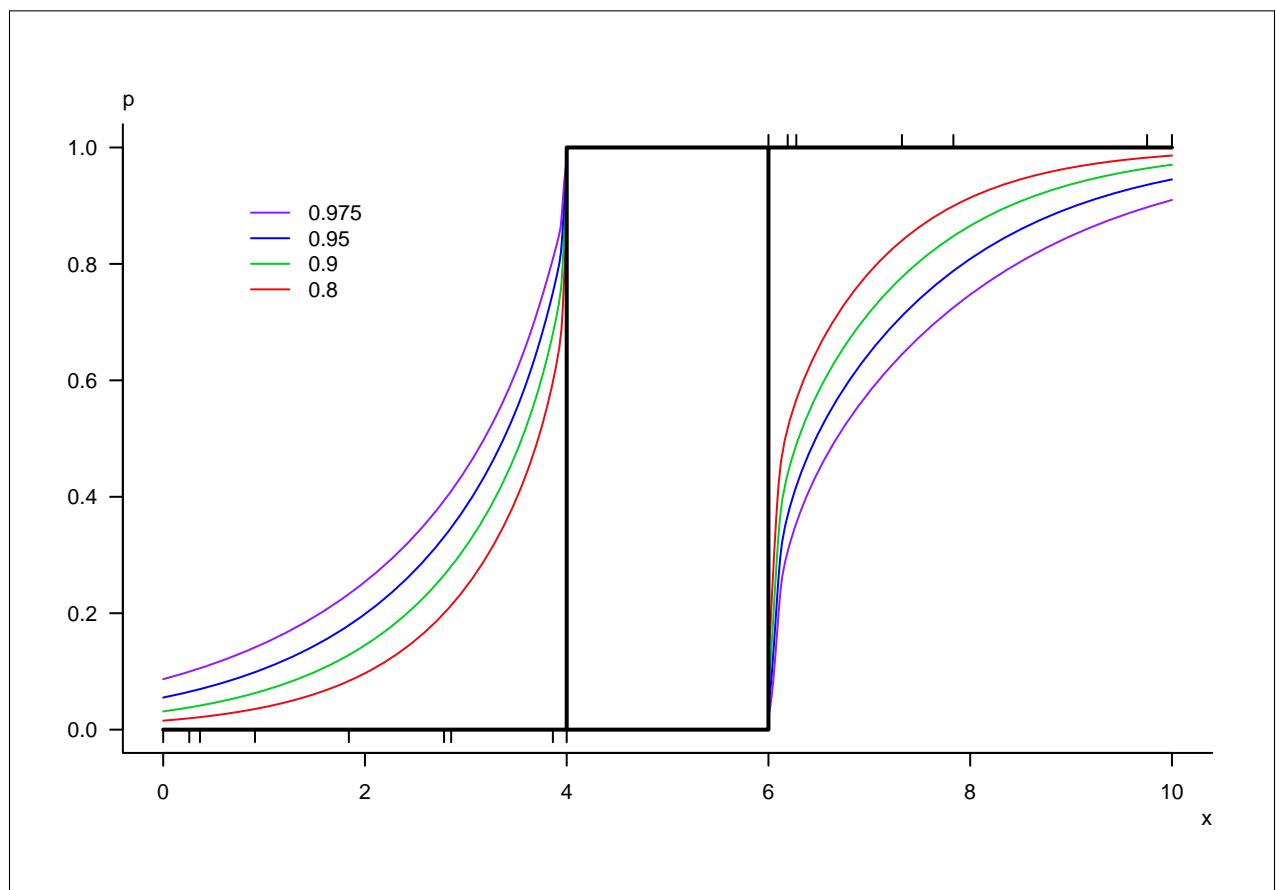


Fig. 17 LR CI bands for gap data, continuity correction

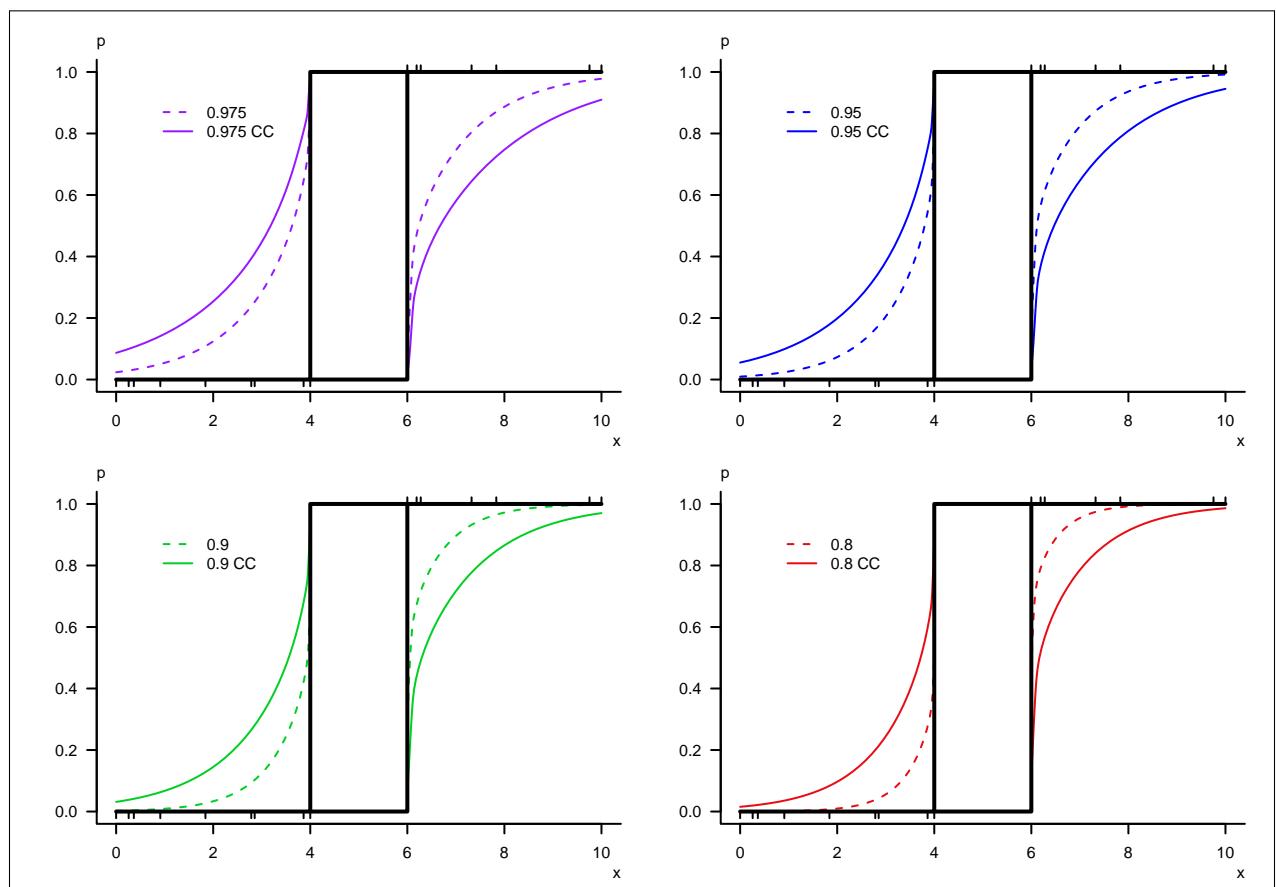


Fig. 18 LR CI bands for gap data, continuity correction (CC) effect

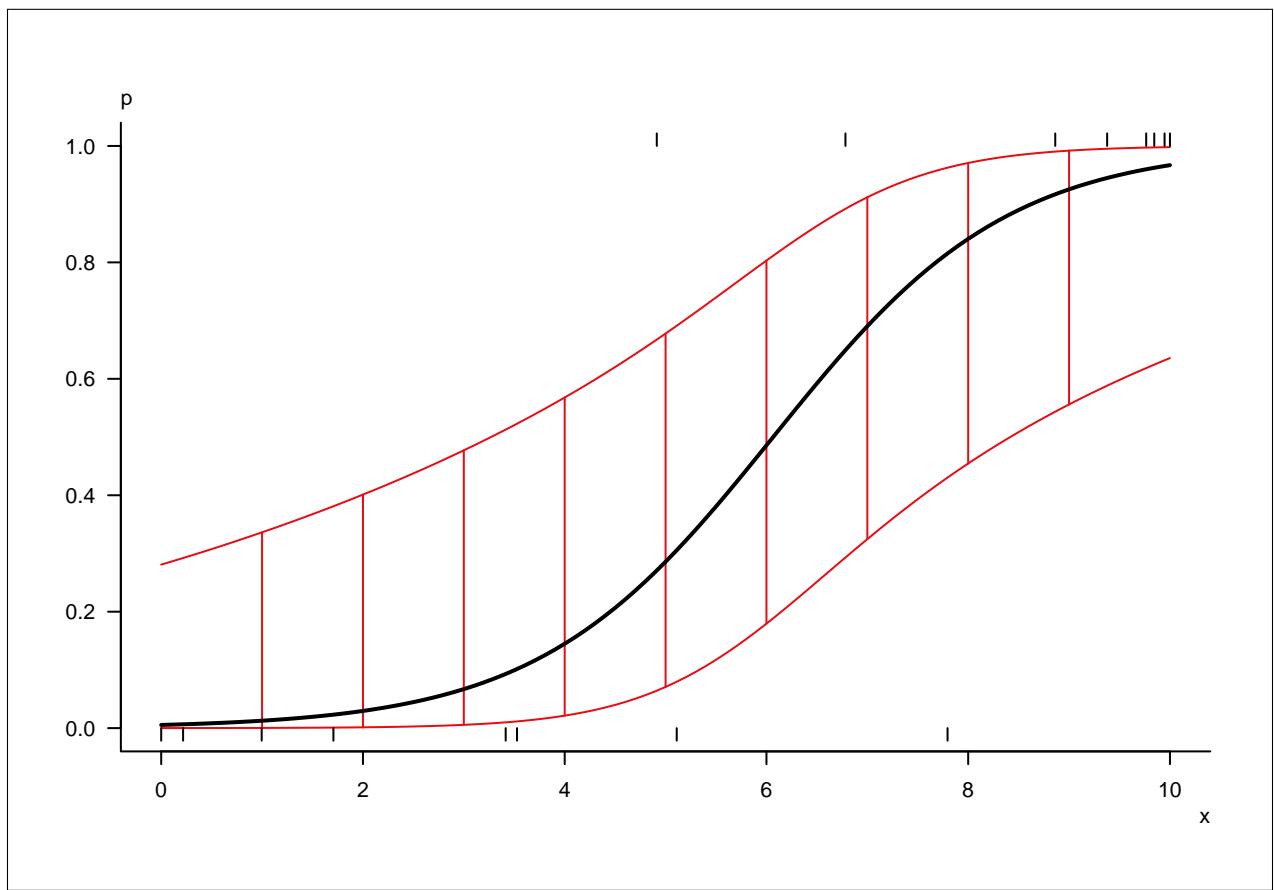


Fig. 19 CR CIs on  $P(x)$  for zmr data

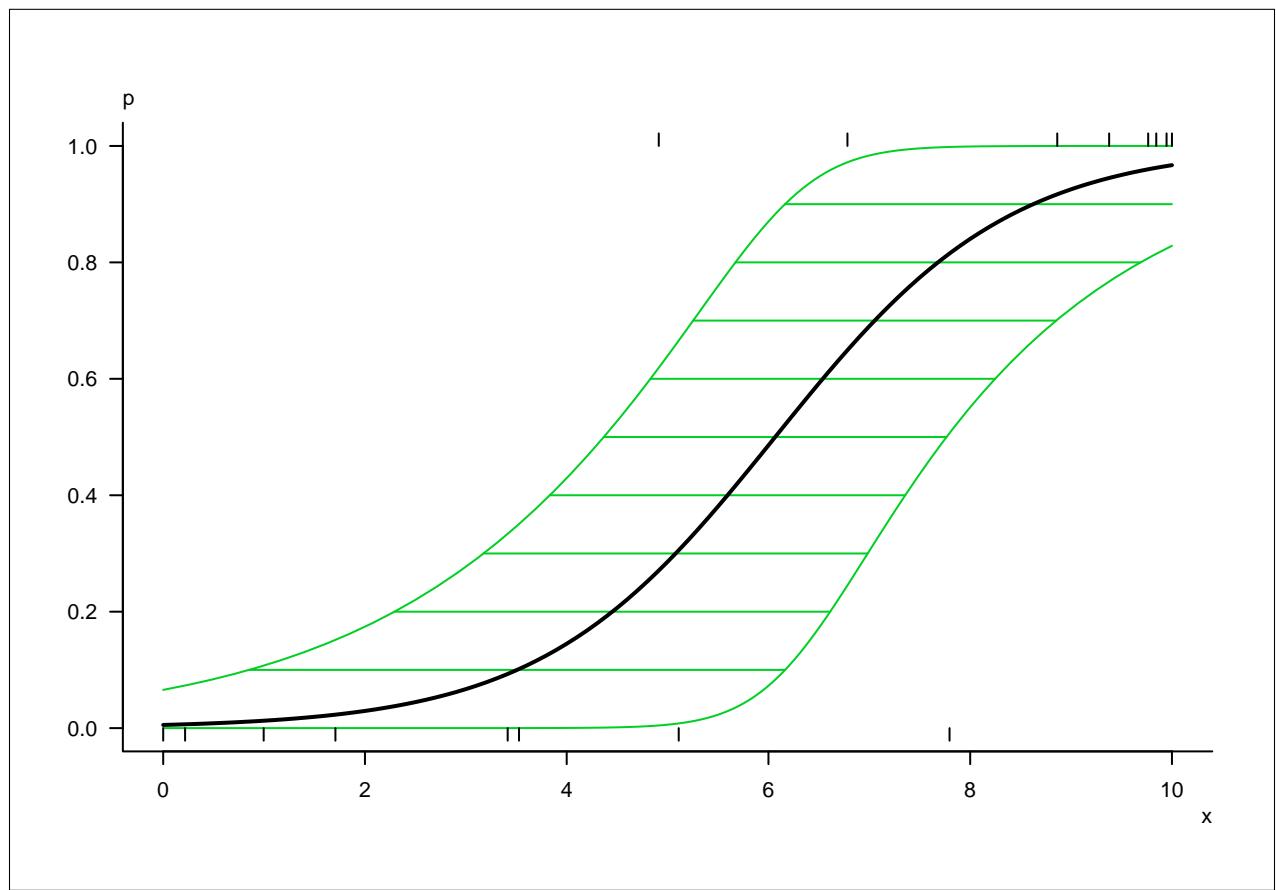


Fig. 20 CR CIs on  $X(p)$  for zmr data

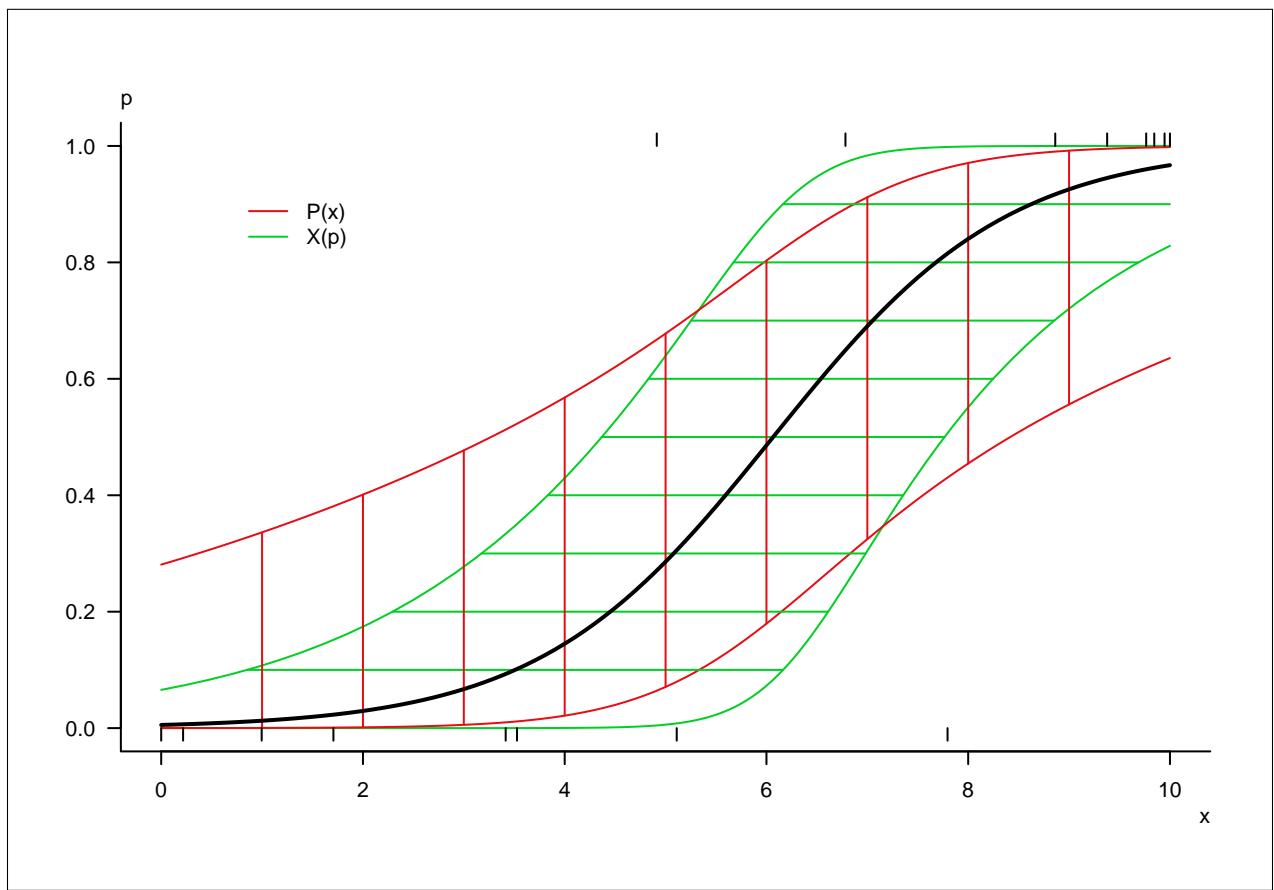


Fig. 21 CR CIs on  $P(x)$  and  $X(p)$  for zmr data

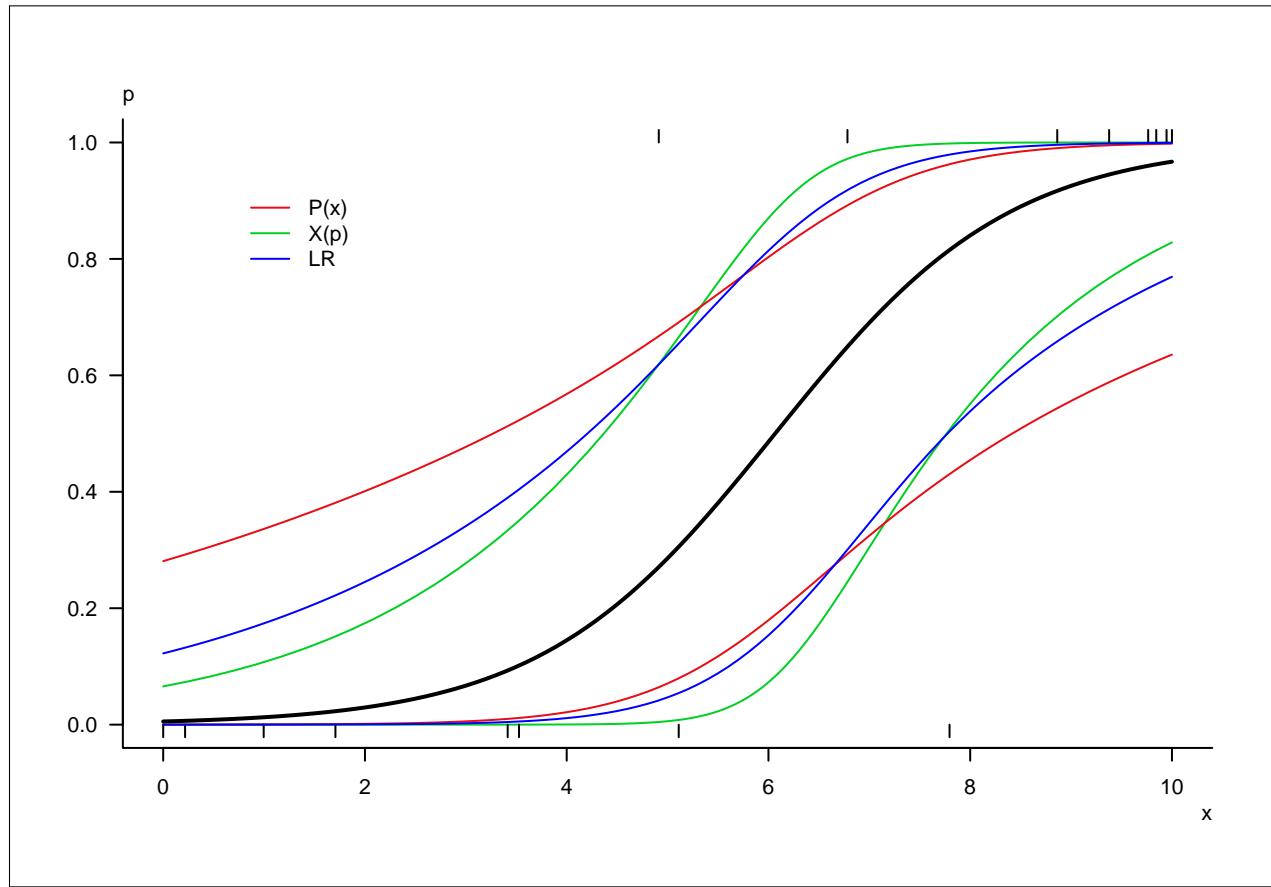


Fig. 22 LR and CR CIs for zmr data

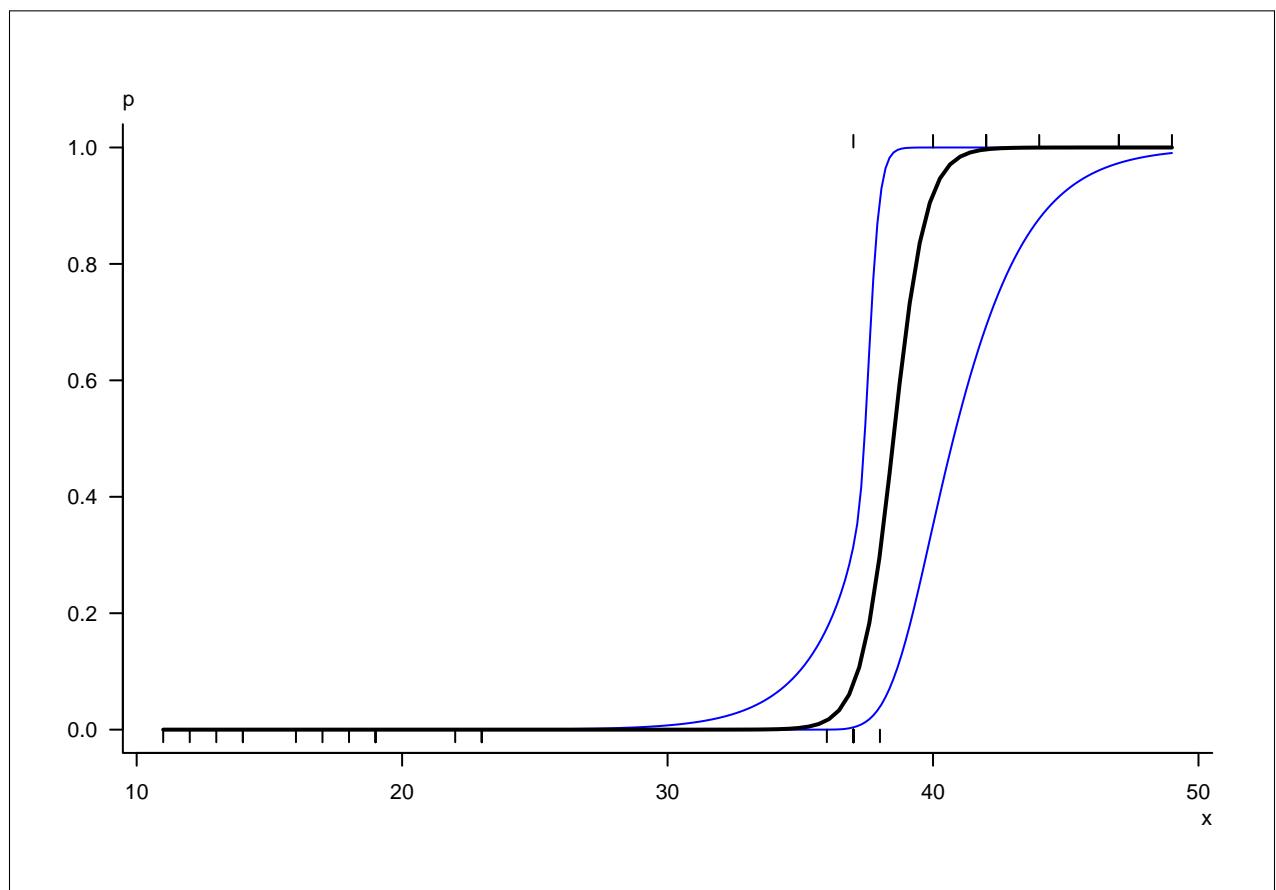


Fig. 23 LR CIs for Webb data

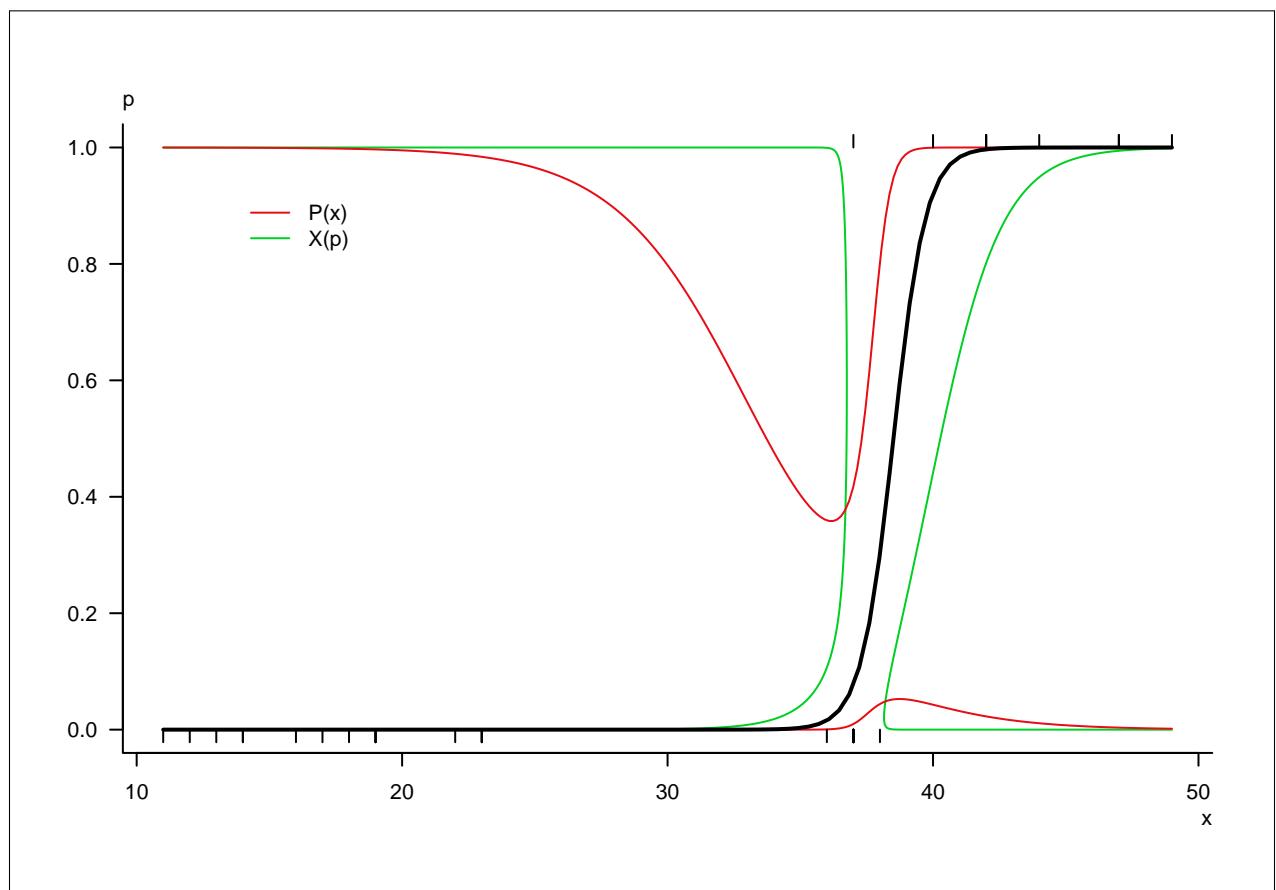


Fig. 24 CR CIs for Webb data

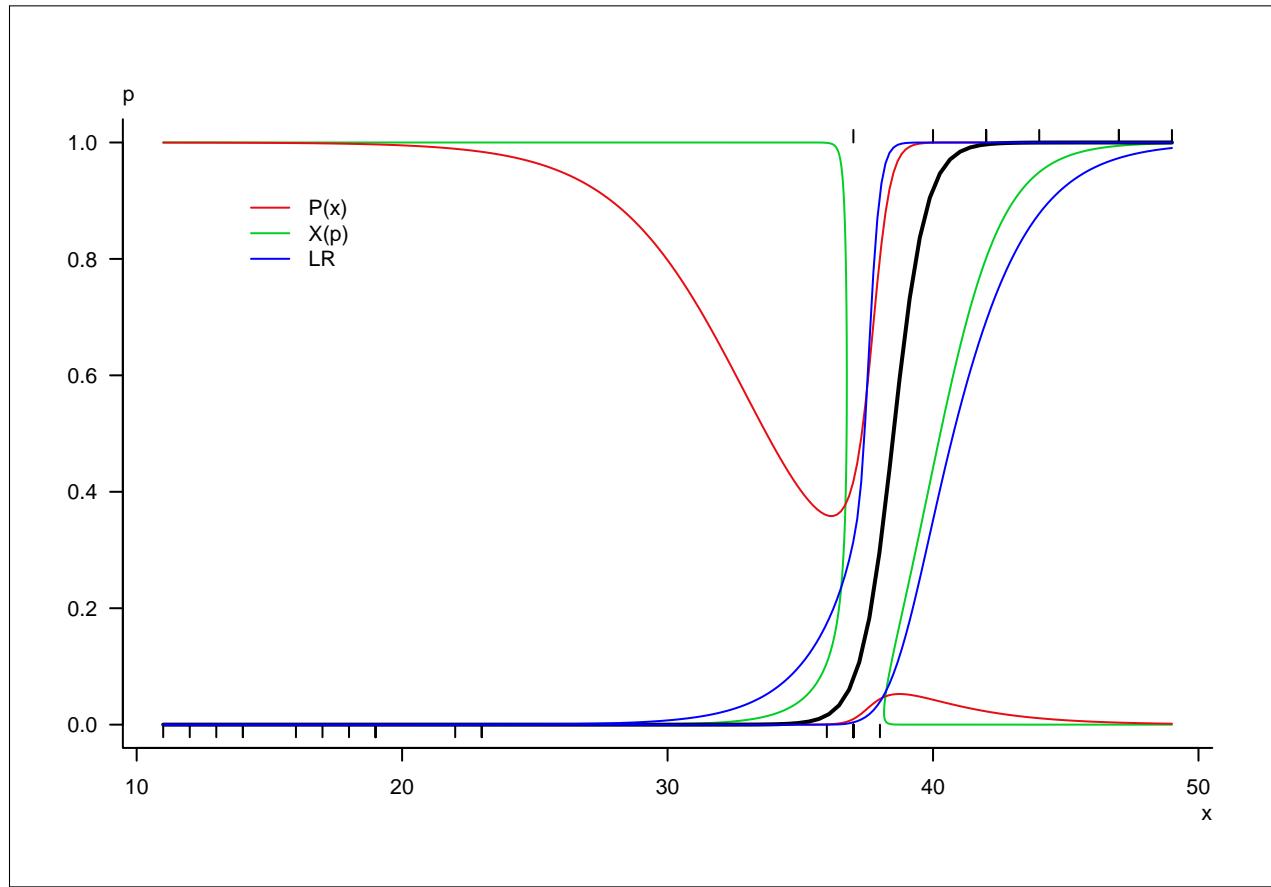


Fig. 25 LR and CR CIs for Webb data

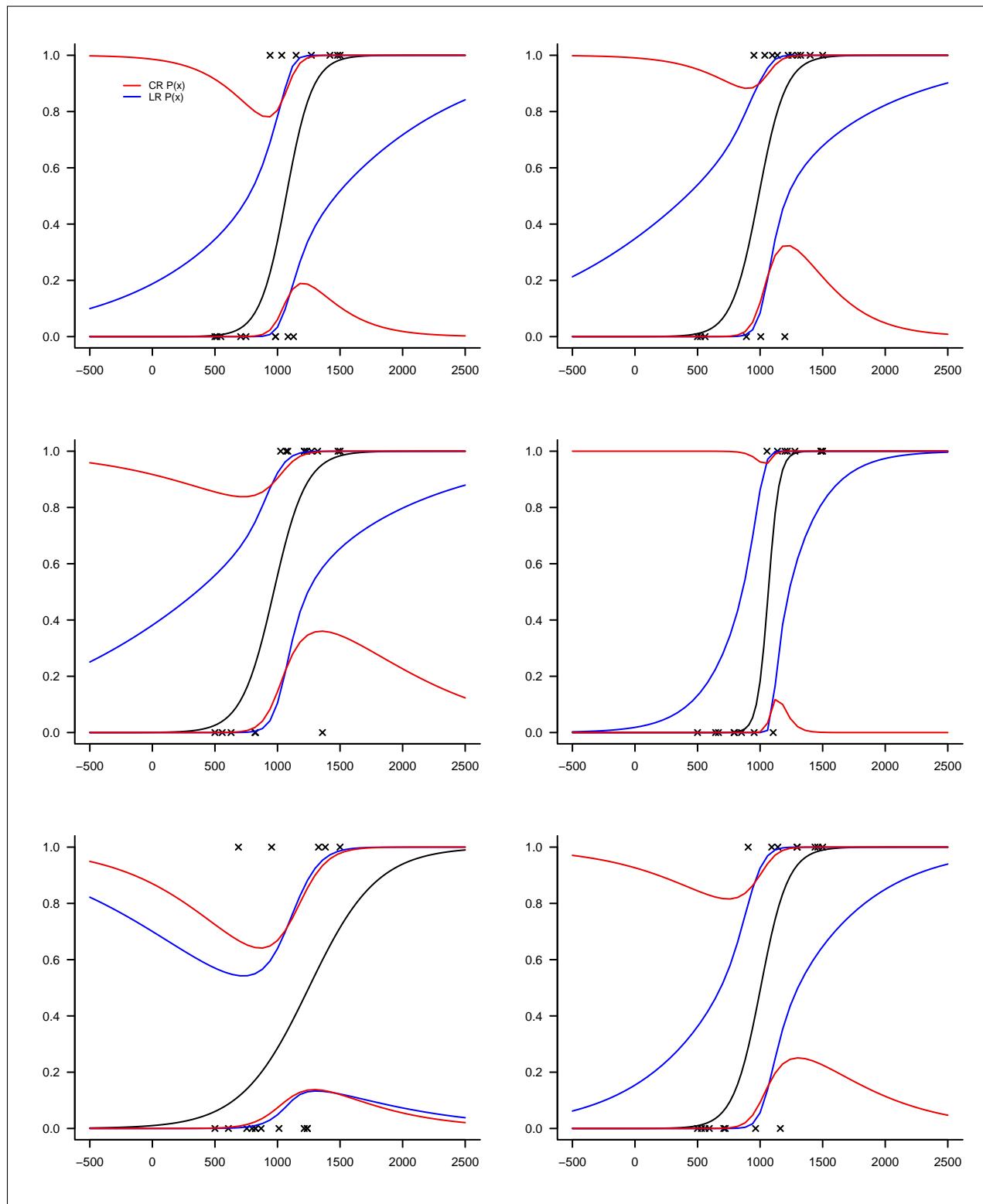


Fig. 26 CR and LR CIs,  $\gamma = 0.99, n = 16$

## 3.4 Two-Sample Inference

### 3.4.1 G Test: the Basic Test

The basic 2-sample G test for the QR model is that both data sets have the same response

$$\begin{aligned} H_0 &: P_1(x) = P_2(x) \\ H_1 &: P_1(x) \neq P_2(x). \end{aligned} \quad (116)$$

We can construct the corresponding GLR test, Eq. 71, evaluate the deviation  $\Delta = D_o - D$ , Eq. 58, and obtain the p-value. The numerator results from common estimation (pooled data) with deviance  $D$ . The denominator corresponds to independent estimation for both sets with deviances  $D_1$  and  $D_2$ . The deviation is

$$\Delta = D - (D_1 + D_2). \quad (117)$$

This works even if either data set (or the pooled data set) has a zmr or not.

Since  $\dim \Theta_0 = 2$  and  $\dim \Theta = 4$ , with

$$\begin{aligned} \Theta_0 &= \{(a, b), (a, b)\} \\ \Theta &= \{(a_1, b_1), (a_2, b_2)\}, \end{aligned} \quad (118)$$

the degrees of freedom is  $r = 2$  for the large-sample approximation.

### 3.4.2 M Test: the Test of X(0.5)

One may wish to test

$$\begin{aligned} H_0 &: X_1(0.5) = X_2(0.5) \\ H_1 &: X_1(0.5) \neq X_2(0.5). \end{aligned} \quad (119)$$

For the ballistic application, this is a test of  $V_{50}$  equality, as  $X(0.5) = V_{50}$ .

If both data sets have a zmr, one may write the models as

$$P_i(x) = G(b_i(x - m_i)) = G\left(\frac{x - m_i}{s_i}\right) \quad (120)$$

for  $i \in \{1, 2\}$ , and the M test is

$$\begin{aligned} H_0 &: m_1 = m_2 \\ H_1 &: m_1 \neq m_2. \end{aligned} \quad (121)$$

The optimal deviation  $\inf_m \Delta(m)$ , where

$$\Delta(m) = (D_1(m) + D_2(m)) - (D_1 + D_2), \quad (122)$$

gives the required optimal common value of  $m = X(0.5)$ . The denominator term  $D_1 + D_2$  corresponds to independent (unrestricted) estimation for both sets with deviances  $D_1$  and  $D_2$ , and the numerator model has common  $m = m_1 = m_2$  and separate  $b_1$  and  $b_2$ .

If both data sets have no zmr, then  $D_1 = D_2 = 0$ .

If both data sets have no zmr and the gap intersection  $Z_1 \cap Z_2 \neq \emptyset$ , then choosing any  $m \in Z_1 \cap Z_2$  yields  $D_1(m) = D_2(m) = 0$ , so  $\Delta = 0$  and  $\hat{p} = 1$ .

If both data sets have no zmr and  $Z_1 \cap Z_2 = \emptyset$ , then for any choice of  $m$ , at least one of  $D_1(m)$  or  $D_2(m)$  is infinite. Therefore  $\Delta = \infty$  and  $\hat{p} = 0$  in this case.

Otherwise, at least 1 data set has a zmr. In this case, for any  $m$  one can evaluate  $D_1(m) + D_2(m)$  by applying Eq. 97 to each data set separately with  $x_o = m$  and  $q_o = 0$ , so  $p_o = 0.5$ , to obtain  $D_1(m)$  and  $D_2(m)$ . The optimal  $m$  minimizes the sum  $D_1(m) + D_2(m)$  and thus gives the numerator term of Eq. 122. The parameter spaces have  $\dim \Theta_0 = 3$  and  $\dim \Theta = 4$ ,

$$\begin{aligned} \Theta_0 &= \{(m, b_1), (m, b_2)\} \\ \Theta &= \{(m_1, b_1), (m_2, b_2)\}, \end{aligned} \quad (123)$$

so  $r = 1$  for the large-sample approximation.

Kinsler and Collins<sup>22</sup> present examples of 2-sample tests.

Figures 27–32 illustrate the application of the G and M tests to pairs of data sets. Figures 27 and 28 show tests with 2 zmr data sets. Figures 29, 30, and 31 show tests with 1 gap and 1 zmr data set. Figure 32 shows tests with 2 gap data sets.

For the G test, the top graph always shows separate response functions (red and green) and the common combined-data response (dashed gray).

For the M test, the bottom left graph shows separate response functions (red and green) and constrained responses (dashed gray) with optimal common  $m = X(0.5)$  and separate  $b$ . The bottom right graph shows the optimization for  $m$  of the p-value  $\hat{p}(m) = 1 - F_{\chi_r^2}(\Delta(m))$ , which achieves its global maximum at the optimal  $m$ . Some care is required in the optimization, as the graph of  $\hat{p}$  may have a local maximum corresponding to the influence of each data set.

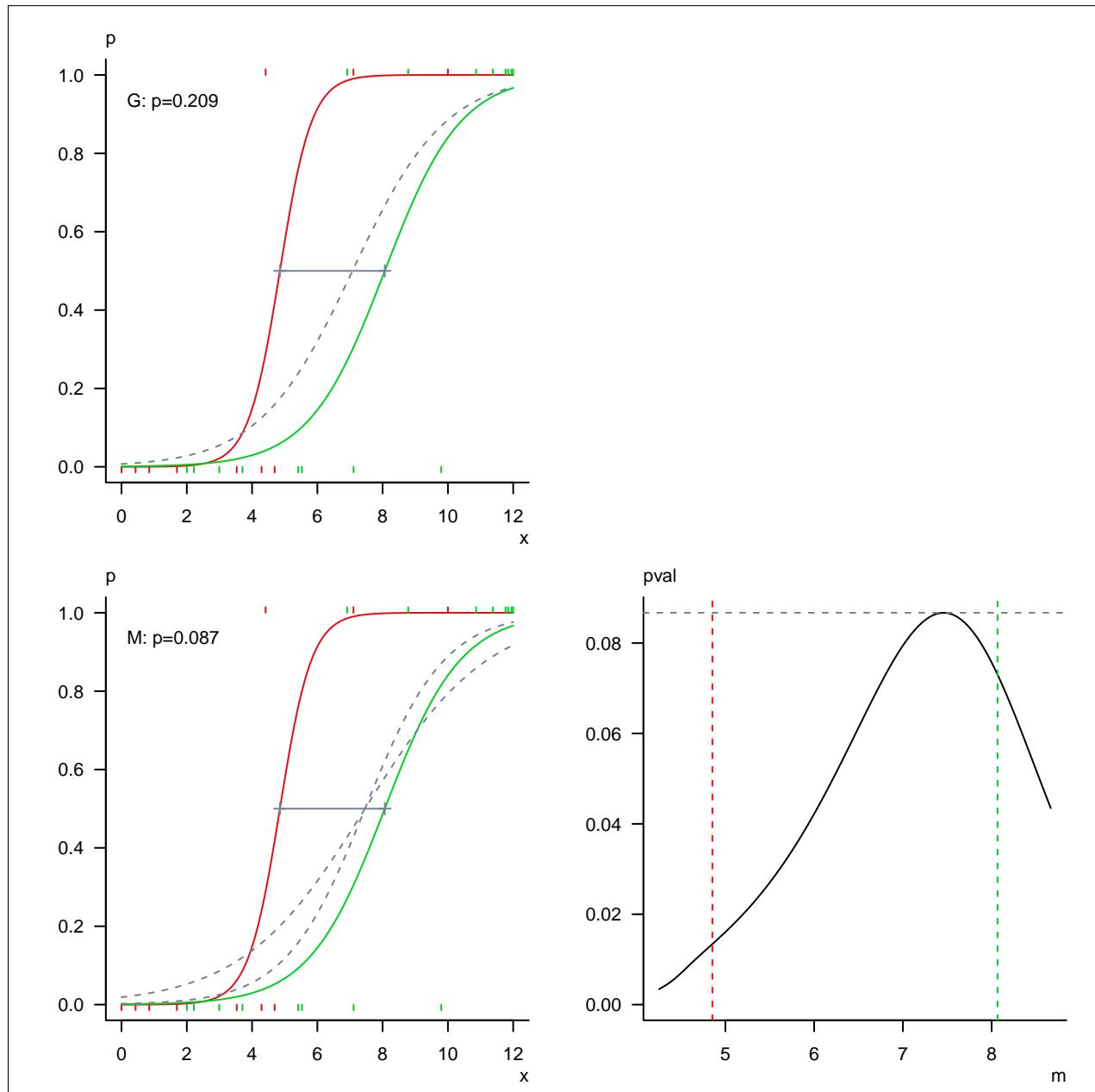


Fig. 27 Two-sample tests, 2 zmr data sets, example 1

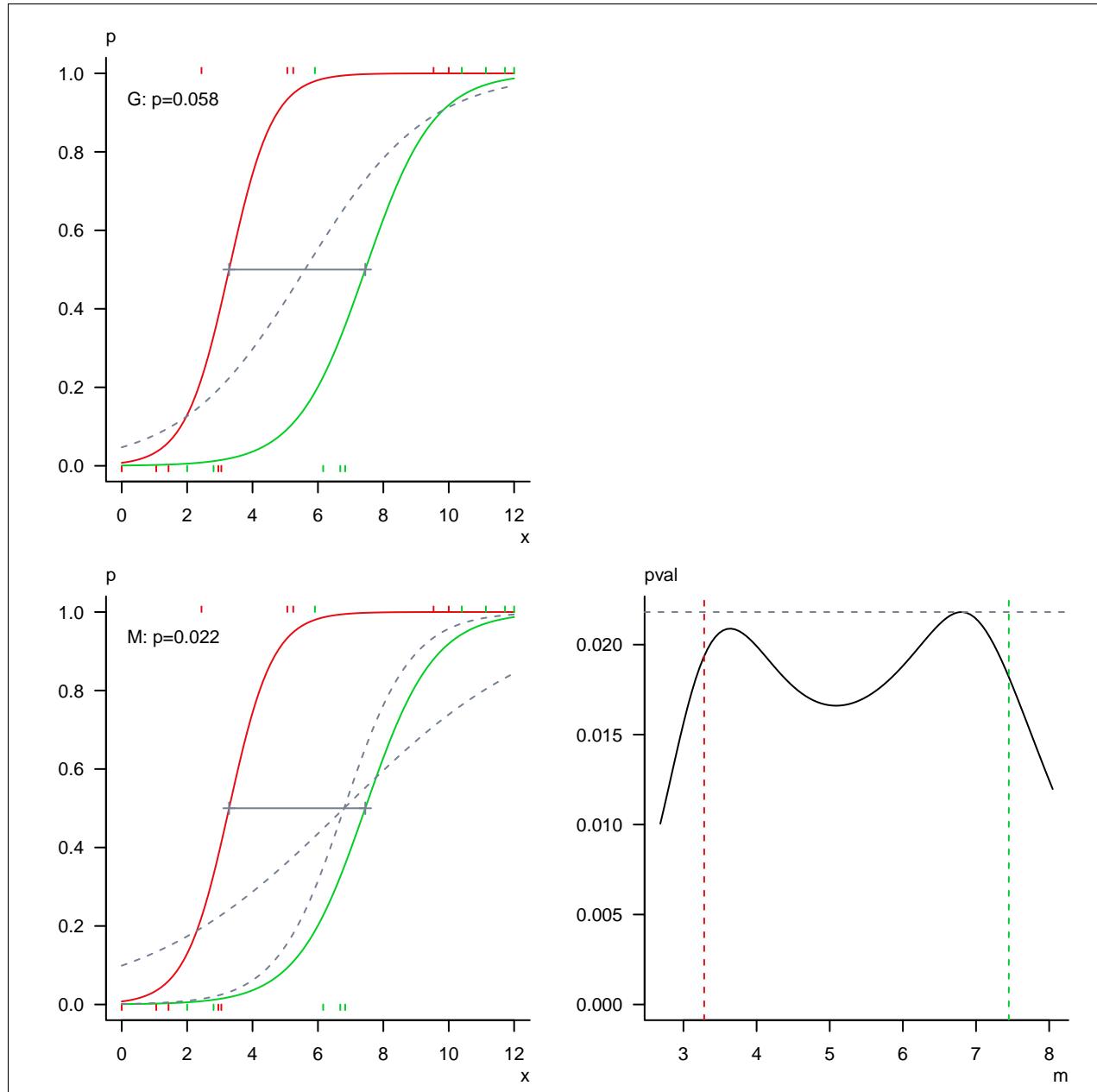


Fig. 28 Two-sample tests, 2 zmr data sets, example 2

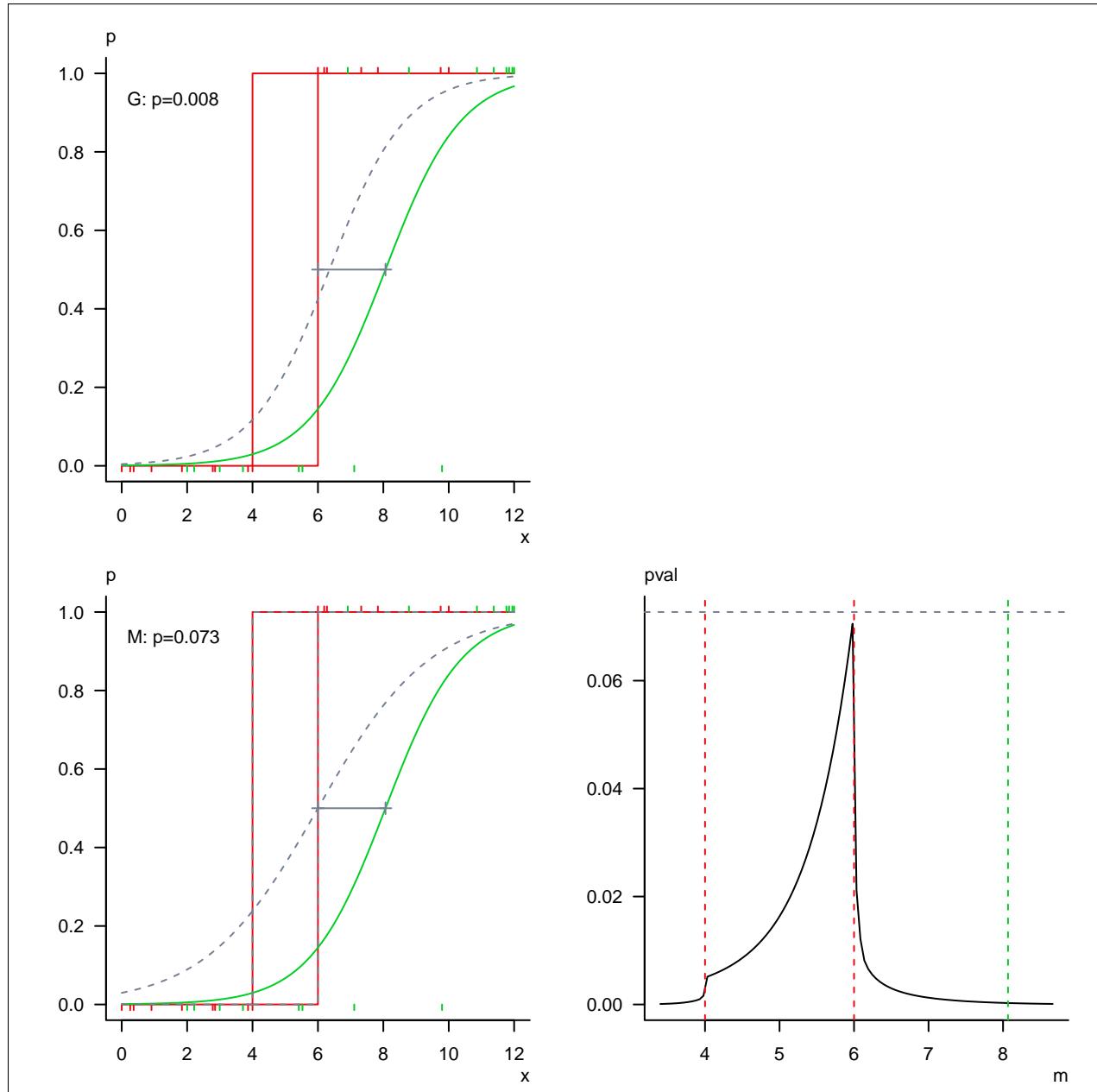


Fig. 29 Two-sample tests, 1 zmr and 1 gap data set, example 1

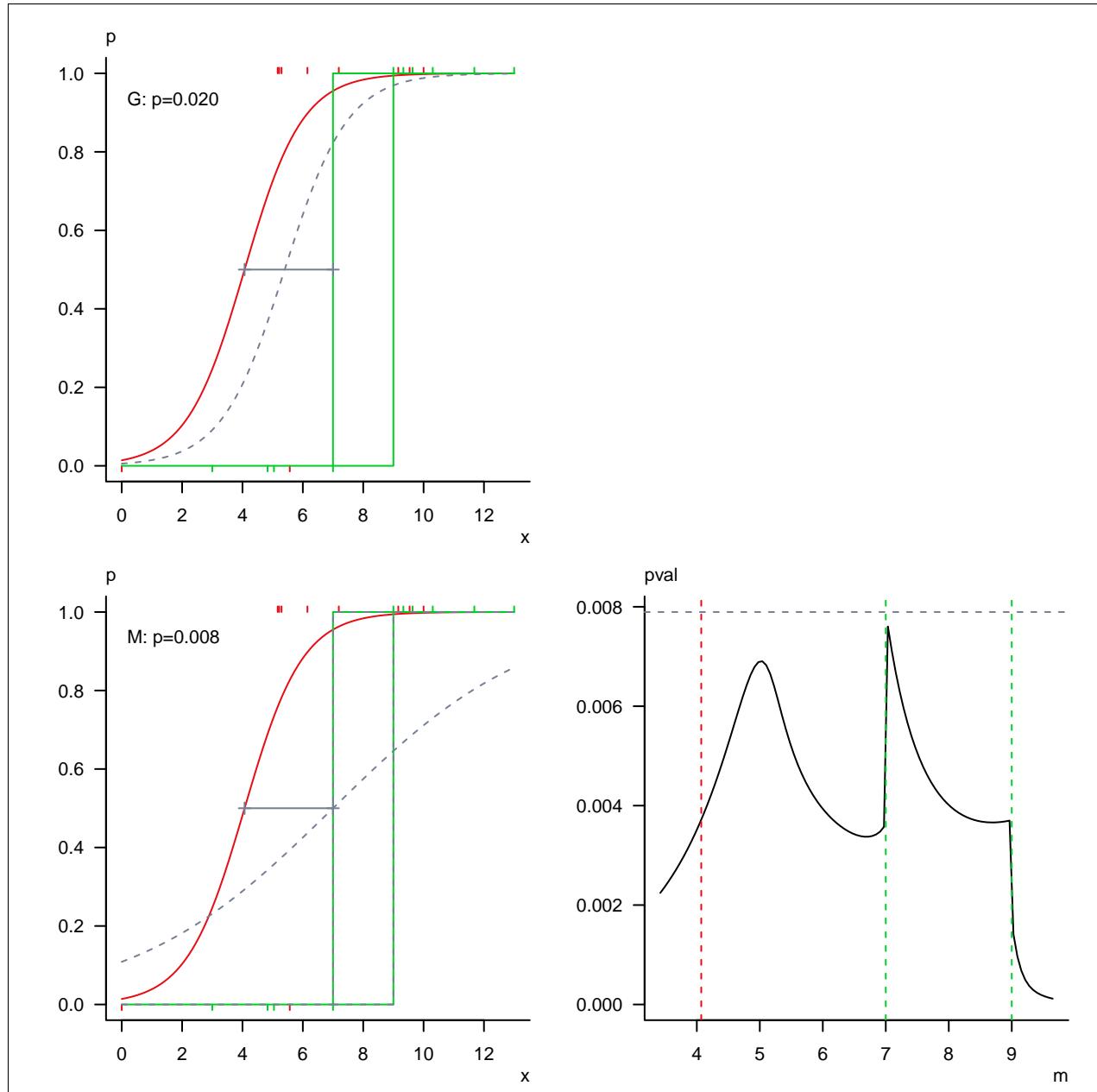


Fig. 30 Two-sample tests, 1 zmr and 1 gap data set, example 2

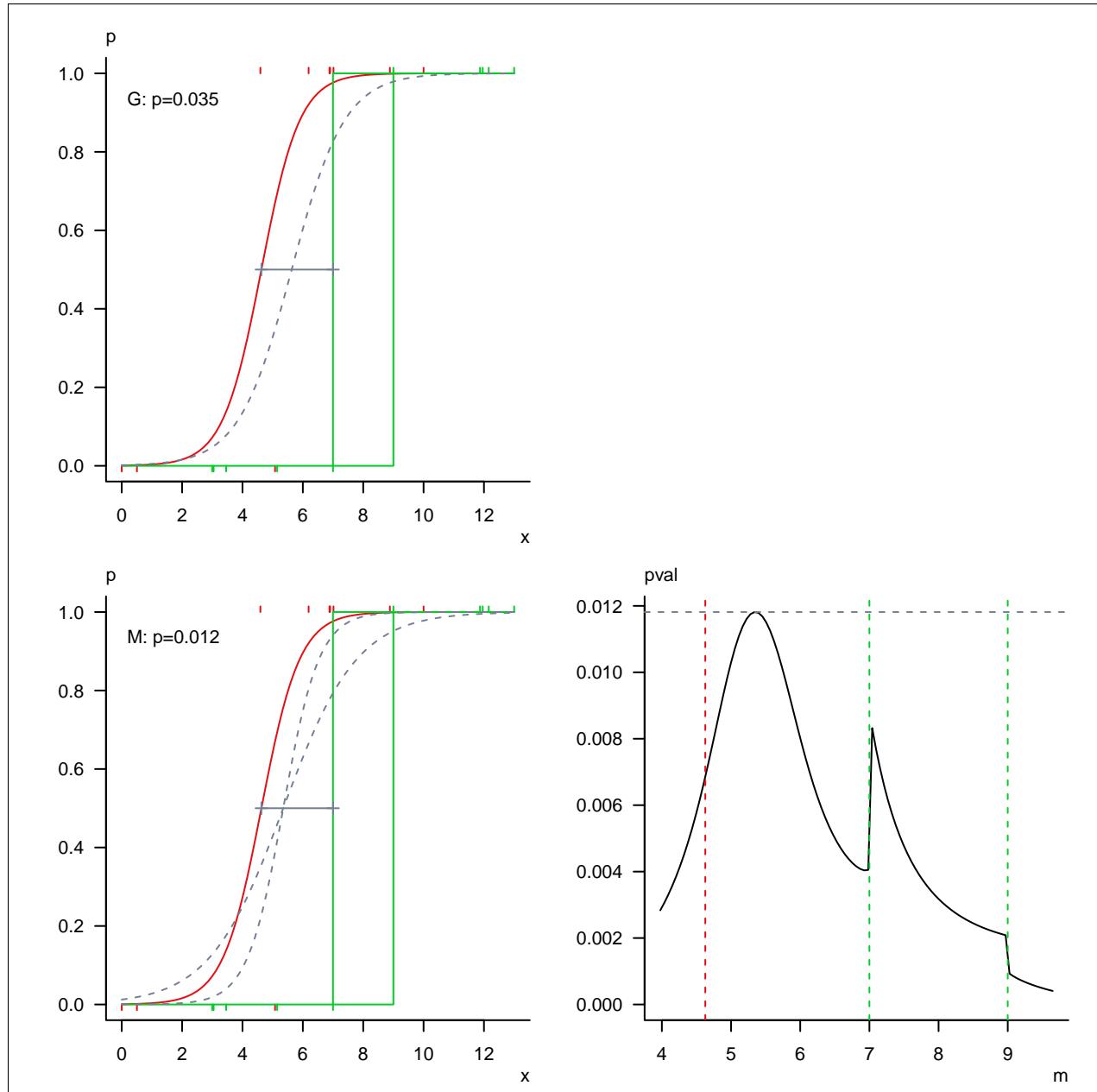


Fig. 31 Two-sample tests, 1 zmr and 1 gap data set, example 3

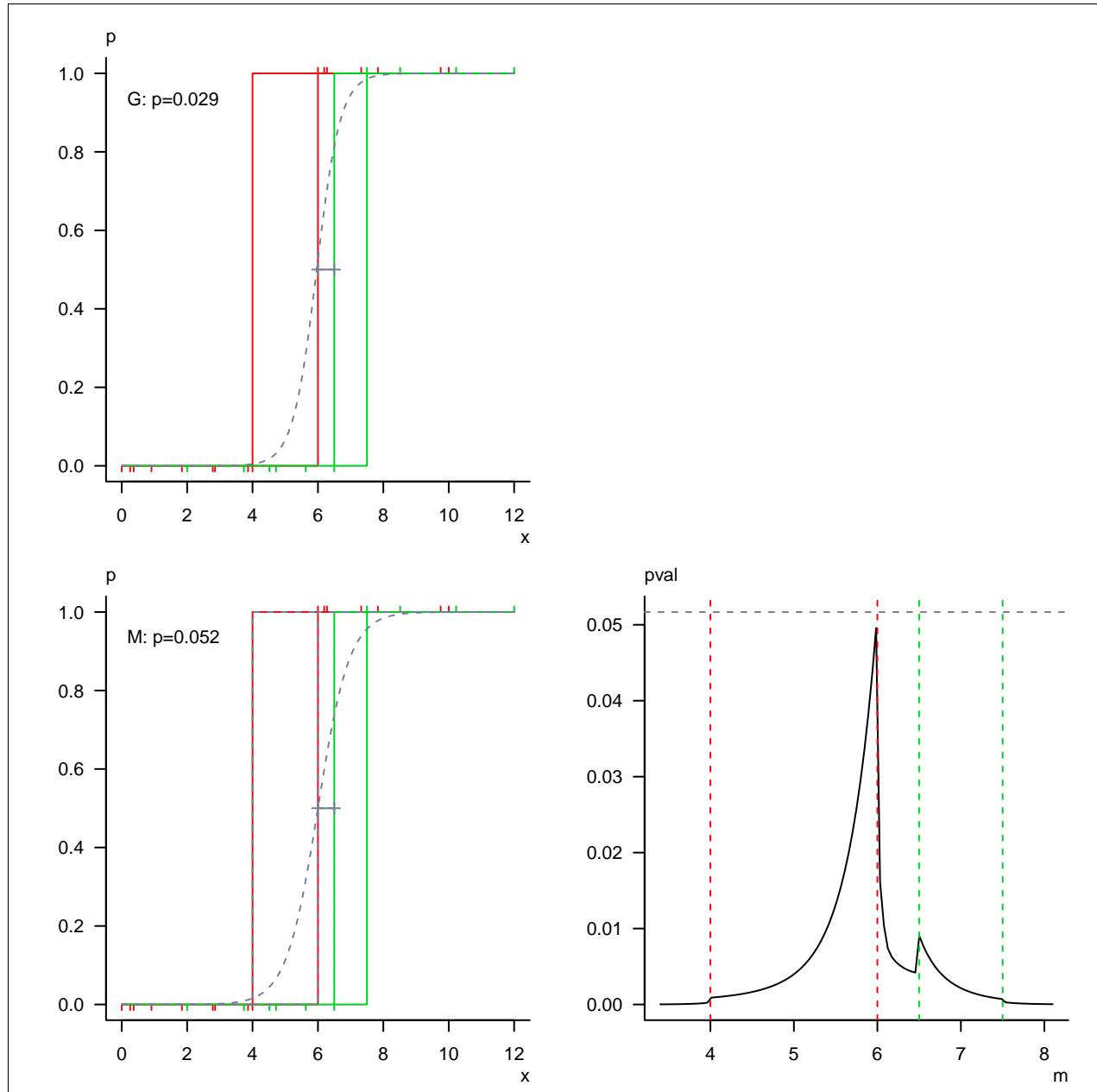


Fig. 32 Two-sample tests, 2 gap data sets

---

## 4. One-Sided Tests

---

Silvapulle<sup>23–25</sup> notes that  $p_1 = p_2/2$  if the statistic  $T$  is in the correct region.

This is easy if  $T \in \mathbb{R}^1$ , when  $T \neq 0$  means  $T < 0$  or  $T > 0$  and there are 2 alternatives.

If  $T = (T_1, \dots, T_k) \in \mathbb{R}^k$ , then  $T = 0$  means  $T_i = 0$  for all  $i$ , and  $T \neq 0$  means at least one component is nonzero. Up to  $k - 1$  components may be 0, the others are positive or negative. So, summing on the number of 0-valued components, the number of alternatives is

$$\sum_{j=0}^{k-1} \binom{k}{j} 2^{k-j} = 3^k - 1. \quad (124)$$

On the other hand, taking  $T \neq 0$  to mean  $T_i \neq 0$  for all  $i$ , there are  $2^k$  alternatives.

The distinction is irrelevant in the continuous case.

---

## 5. Computation

---

Suppose we have a data set  $\{(x_1, y_1), \dots, (x_n, y_n)\}$  of size  $n$ . Let  $n_0$  be the number of points with  $y = 0$  and  $n_1$  be the number of points with  $y = 1$ .

The link is  $G(z)$  and its derivative is

$$g(z) = \frac{d}{dz} G(z). \quad (125)$$

The upper-tail link is  $H(z) = 1 - G(z)$  and its derivative is

$$h(z) = \frac{d}{dz} H(z) = -g(z). \quad (126)$$

The likelihood is

$$L = \prod_{i=1}^n G(z_i)^{y_i} H(z_i)^{1-y_i} \quad (127)$$

and its logarithm is  $\ell = \log L$

$$\ell = \sum_{i=1}^n (y_i \log G(z_i) + (1 - y_i) \log H(z_i)) = \sum_{y_i=1} \log G(z_i) + \sum_{y_i=0} \log H(z_i) \quad (128)$$

and its derivative

$$\frac{d}{du} \ell = \sum_{y_i=1} \frac{dz_i}{du} \cdot \frac{g(z_i(u))}{G(z_i(u))} + \sum_{y_i=0} \frac{dz_i}{du} \cdot \frac{h(z_i(u))}{H(z_i(u))}. \quad (129)$$

## 5.1 Full Model

Collins<sup>11</sup> describes the full model  $z = a + bx$ ,

$$P(x) = G(a + bx). \quad (130)$$

## 5.2 Null Model: $\mathbf{b} = \mathbf{0}$

The null model in R (and S) nomenclature is the null model for the particular test

$$\begin{aligned} H_0 &: b = 0 \\ H_1 &: b \neq 0 \end{aligned} \quad (131)$$

(slope  $b = 0$ ) in the model  $P(x) = G(a + bx)$ . In other words,

$$P(x) = G(a) \quad (132)$$

so  $z = a$ , and with

$$\frac{d}{da} \ell = n_1 \frac{g(a)}{G(a)} - n_0 \frac{g(a)}{1 - G(a)} = 0 \quad (133)$$

it follows that  $n_1(1 - G(a)) = n_0G(a)$ , so  $G(a) = n_1/(n_0 + n_1)$  and

$$a = Q\left(\frac{n_1}{n}\right). \quad (134)$$

## 5.3 Restricted Model: $\mathbf{G}(\mathbf{x}_0) = \mathbf{p}_0$

To estimate  $b$  in the restricted model

$$P(x) = G(q_o + b(x - x_o)) \quad (135)$$

without loss of generality (using  $x - x_o$  for  $x$  and 0 for  $x_o$  and  $a$  for  $q_o$ ), one may work with

$$P(x) = G(a + bx). \quad (136)$$

So let  $z = a + bx$ , and maximize the likelihood by setting

$$\frac{d}{db} \ell(b) = 0. \quad (137)$$

For convenience write  $G_i = G(z_i)$  and  $H_i = H(z_i)$ , so that

$$\ell(b) = \sum_{y_i=1} \log G_i + \sum_{y_i=0} \log H_i. \quad (138)$$

Then

$$\begin{aligned} \frac{d}{db} \ell(b) &= \sum_{y_i=1} x_i \cdot \frac{g_i}{G_i} + \sum_{y_i=0} x_i \cdot \frac{h_i}{H_i} \\ &= \sum_{y_i=1} x_i \cdot \frac{g(a+bx_i)}{G(a+bx_i)} + \sum_{y_i=0} x_i \cdot \frac{h(a+bx_i)}{H(a+bx_i)}. \end{aligned} \quad (139)$$

Numerical solution of  $d\ell/db = 0$  is easy, and the root maximizes  $\ell$ .

---

## 6. Conclusions

---

Ballistic limit testing and acceptance criteria should be expressed in terms of statistical tests that account for quantification of risk and uncertainty. The LR methodology developed in this report satisfies the requirement for statistical decision support of ballistic limit analyses and applies equally to experiments with or without a zmr. The level of computational detail required is easily achievable in modern computing environments. Any interested reader may contact the author for assistance. Discussion and commentary are also welcome.

---

## 7. References

---

1. Golub A, Grubbs F. On estimating ballistic limit and its precision. Aberdeen Proving Ground (MD): Ballistic Research Laboratory (US); 1950 Mar. Report No.: BRL-TN-151.
2. Grubbs FE. Selected topics in experimental statistics with Army applications, Volume 706, Issue 103 of DARCOM pamphlet. Alexandria (VA): Army Materiel Development and Readiness Command; 1983 Dec.
3. McKaig AE, Thomas J. Maximum likelihood program for sequential testing documentation. Aberdeen Proving Ground (MD): Ballistic Research Laboratory (US); 1983 Mar. Report No.: ARBRL-TR-02481.
4. DiDonato AR, Jarnagin MP Jr. Use of the maximum likelihood method under quantal responses for estimating the parameters of a normal distribution and its application to an armor penetration problem. Dahlgren (VA): Naval Weapons Laboratory (US); 1972 Mar. Technical Report No.: TR-2846.
5. Test operations procedure: ballistic tests of armor materials. Aberdeen Proving Ground (MD): US Army Test and Evaluation Command (US). 1984 Feb. TOP 2-2-710, ADA137873.
6. Hagan J. Computer program for analysis of sensitivity data following a normal distribution. Aberdeen Proving Ground (MD); 1962 Aug. Analytical Laboratory Report No.: 62-AL-154. Also available at: <http://www.dtic.mil/docs/citations/AD0661636>.
7. MIL-STD-662. Department of defense test method standard, V<sub>50</sub> ballistic test for armor. Aberdeen Proving Ground (MD): Army Research Laboratory (US); 1997 Dec.
8. MIL-DTL-46100E. Detail specification, armor plate, steel, wrought, high hardness. Aberdeen Proving Ground (MD): Army Research Laboratory (US); 2008 Jul.
9. MIL-DTL-32332. Detail specification, armor plate, steel, wrought, ultra high hardness. Aberdeen Proving Ground (MD): Army Research Laboratory (US); 2009 Jul.
10. Hauck WW, Donner A. Wald's test as applied to hypotheses in logit analysis. Journal of the American Statistical Association. 1977;72(360):851–853.
11. Collins J. Quantal response: practical sensitivity testing. Aberdeen Proving Ground (MD): Army Research Laboratory (US); 2012 Jun. Report No.: ARL-TR-6022. Also available at [http://www.arl.army.mil/www/default.cfm?technical\\_report=6491](http://www.arl.army.mil/www/default.cfm?technical_report=6491).

12. Webb DW. Quantal response analysis in the absence of a zone of mixed results using data augmentation. Aberdeen Proving Ground (MD): Ballistic Research Laboratory (US); 2008 Sep. Report No.: ARL-TR-4604. Also available at [http://www.arl.army.mil/www/default.cfm?technical\\_report=1686](http://www.arl.army.mil/www/default.cfm?technical_report=1686).
13. Neyer Software. Cincinnati (OH). Neyer Software; [accessed 2014 Jul 29; 2014 Jul 29]. <http://neyersoftware.com>.
14. Stuart A, Ord JK. Kendall's advanced theory of statistics: classical inference and relationship. New York: Oxford University Press; 1991.
15. Wilks SS. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Annals of Mathematical Statistics*. 1938;9:60–62.
16. The R Project for Statistical Computing. Wien (AT): The R Foundation for Statistical Computing; [accessed 2014 Jul 29; 2014 May 22]. <http://www.r-project.org>.
17. Becker RA, Chambers JM, Wilks AR. The new S language. London: Chapman & Hall; 1988.
18. Wolfram Mathematica: Definitive System for Modern Technical Computing. Champaign (IL): Wolfram; [accessed 2014 Jul 29; 2014 Jul 29]. <http://www.wolfram.com/mathematica>.
19. MATLAB - The Language of Technical Computing - B. Natick (MA): MathWorks; [accessed 2014 Jul 29; 2014 Jul 29]. <http://www.mathworks.com/products/matlab>.
20. JMP 11. Cary (NC): SAS Institute; [accessed 2014 Jul 29; 2014 Jul 29]. <http://www.jmp.com/software/jmp>.
21. Kinsler R, Collins J. V<sub>50</sub> analysis of Benteler steel plates. Aberdeen Proving Ground (MD): Army Research Laboratory (US); 2014 Sep. Report No.: ARL-SR-0297.
22. Kinsler R, Collins J. Comparison of V<sub>50</sub> shot placement on final outcome. In: PASS 2012 Conference Proceedings; 2012 Sep 17–21 ; Nuremberg, Germany. Personal Armour Systems Symposium; 2012 Sep 17–21.
23. Silvapulle MJ. On likelihood ratio tests of one-sided hypotheses in generalized linear models with canonical likns. Chapel Hill (NC): University of North Carolina, Department of Biostatistics; 1990 May. Institute of Statistics Mimeo Series No.: 1877.

24. Silvapulle MJ. Robust Wald-type tests of one-sided hypotheses in the linear model. Chapel Hill (NC): University of North Carolina, Department of Biostatistics; 1990 May. Institute of Statistics Mimeo Series No.: 1878.
25. Silvapulle MJ. Robust tests of inequality constraints and one-sided hypotheses in the linear model. Chapel Hill (NC): University of North Carolina, Department of Biostatistics; 1990 May. Institute of Statistics Mimeo Series No.: 1879.

---

## List of Symbols, Abbreviations, and Acronyms

---

BL(P)	Ballistic limit, protection criteria
cdf	cumulative distribution function
CI	confidence interval
CR	Cramér-Rao
DTL	Detail Specification
GLM	Generalized Linear Model
GLR	generalized likelihood ratio
LR	likelihood ratio
MLE	maximum likelihood estimation
QR	quantal response
SLR	simple likelihood ratio
TOP	Test Operations Procedure
USADTC	US Army Developmental Test Command
zmr	zone of mixed results

1 (PDF)	DEFENSE TECHNICAL INFORMATION CTR DTIC OCA	8 (PDF)	ATEC AEC N DUNN Z ZIMMER DTC C ANDRES W BOUGHERS B GILICH S MEISELWITZ TEDT AT SLE L LOMBARDO TEAE IS E D WEBB
2 (PDF)	DIRECTOR US ARMY RESEARCH LAB RDRL CIO LL IMAL HRA MAIL & RECORDS MGMT		
1 (PDF)	GOVT PRINTG OFC A MALHOTRA		
1 (HC)	DIR US ARMY EVALUATION CTR HQ TEAE SV P A THOMPSON 2202 ABERDEEN BLVD 2ND FL APG MD 21005-5001	1 (PDF)	US ACE EDRC N BOONE
3 (2HC 1 PDF)	DIR USARL RDRL SL J BEILFUSS (HC) P TANENBAUM (HC) RDRL SLE R FLORES	3 (PDF)	US NAVY NSWCDD D DICKINSON NAVAIR R HAGAN K RUSSELL
19 (PDF)	DIR USARL RDRL CII C B BODT RDRL DPW T HOLDREN RDRL SLB R BOWEN RDRL SLB A G MANNIX RDRL SLB D J COLLINS J EDWARDS R GROTE L MOSS E SNYDER RDRL SLB E M MAHAFFEY RDRL SLB G P MERGLER RDRL SLB S M PERRY R SAUCIER RDRL SLB W S SNEAD RDRL WML A A THOMPSON RDRL WMP E D GALLARDY E HORWATH M KORNECKI P SWOBODA		